

72798

mg

184368

**STIC-Biotech/ChemLib**

---

**From:** Portner, Ginny  
**Sent:** Wednesday, April 05, 2006 2:30 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 10/768,093

please search SEQ Id No 9 and 4 with respect to polypeptide/amino acid sequences.

Ginny Portner  
Remsen Building  
Art Unit 1645  
Room E03, B02; Mail Box 3C18  
(571) 272-0862

RECEIVED  
APR - 5 2006  
.../CHEM. LIB.  
(STIC)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: April 10, 2006, 10:13:49 ; Search time 230 Seconds  
(without alignments)

417.182 Million cell updates/sec

Title: US-10-768-093-9  
Perfect score: 699  
Sequence: 1 RTIATKPNFVSTTISKSPF.....EXKISGIVNDQVWVGYYVN 136

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: UniProt\_05.80.\*  
1: UniProt\_sprot.\*  
2: UniProt\_trembl.\*

'Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	699	100.0	154	1 F6A2_ECOLI	P33509 escherichia
2	655	93.7	154	1 F6A1_ECOLI	P33508 escherichia
3	120.5	17.2	167	1 F6B1_ECOLI	P33510 escherichia
4	113.5	16.2	167	1 F6B2_ECOLI	P33511 escherichia
5	92	13.2	440	2 05AUS6_DICDI	05AUS6 dictyosteli
6	87	12.4	918	2 072P30_LEPIC	072P30 leptospira
7	87	12.4	918	2 08F7D7_LEPIN	08F7D7 leptospira
8	83	11.9	749	2 081A13_BACCR	081A13 bacillus ce
9	83	11.9	12268	2 08M008_CAEEL	08M008 caenorhabdi
10	83	11.9	13100	2 009165_CAEEL	009165 caenorhabdi
11	82	11.7	248	2 06H0E5_BACAN	06H0E5 bacillus an
12	82	11.7	248	2 0630P3_BACCR	0630P3 bacillus ce
13	82	11.7	248	2 06HASE_BACHK	06HASE bacillus th
14	82	11.7	264	2 081J14_BACAN	081J14 bacillus an
15	82	11.7	268	2 04MG61_BACCE	04MG61 bacillus ce
16	82	11.7	279	2 084214_9BACT	084214 uncultured
17	82	11.7	287	2 072X32_BACCI	072X32 bacillus ce
18	82	11.7	524	2 06CDU2_YARLI	06CDU2 yarrowia li
19	81.5	11.7	449	2 087125_VIBPA	087125 vibrio para
20	81.5	11.7	454	2 0897U6_CLOTE	0897U6 clostridium
21	81.5	11.7	610	2 05WV11_LEGPI	05WV11 legionella
22	81.5	11.7	610	2 05X443_LEGPA	05X443 legionella
23	81.5	11.7	610	2 05ZUD2_LEGPA	05ZUD2 legionella
24	81	11.6	318	2 04K1L8_STRPN	04K1L8 streptococc
25	81	11.6	318	2 04K1O5_STRPN	04K1O5 streptococc
26	81	11.6	595	2 08YU11_ANASP	08YU11 anabena sp
27	81	11.6	1047	2 09RB35_9SPHI	09RB35 cytophaga s
28	81	11.6	1838	2 088207_MOUSE	088207 mus musculu
29	80.5	11.5	360	1 AROC_HAEDU	07V142 haemophilus
30	80.5	11.5	1303	2 04WIH6_ASPFU	04WIH6 aspergillus
31	80	11.4	251	2 0814R6_BACCR	0814R6 bacillus ce

32	80	11.4	279	2 084219_9BACT	084219 uncultured
33	80	11.4	316	2 0523K0_MAGCR	0523K0 magnetotriche
34	80	11.4	440	2 08A503_BACTN	08A503 bacteroides
35	80	11.4	1936	2 04ODF6_LEIMA	04ODF6 leishmania
36	79.5	11.4	385	2 07RRK4_PLAIO	07RRK4 plasmodium
37	79.5	11.4	459	2 08D3R5_VIBRU	08D3R5 vibrio vuln
38	79.5	11.4	826	2 08D3X7_STRAS	08D3X7 streptococc
39	79	11.3	1372	2 05PB95_ANAMM	05PB95 anaplasma m
40	79	11.3	1840	2 09J103_RAT	09J103 rattus norv
41	78.5	11.2	279	2 05K2E9_9BACT	05K2E9 uncultured
42	78.5	11.2	279	2 05K2F1_9BACT	05K2F1 uncultured
43	78.5	11.2	279	2 05K2F7_9BACT	05K2F7 uncultured
44	78.5	11.2	279	2 05K2F9_9BACT	05K2F9 uncultured
45	78.5	11.2	279	2 05K2H9_9BACT	05K2H9 uncultured

Wolfe et al. Submitted 1/94

RESULT 2	Query Match	Best local similarity	Matches	Conservative	Mismatches	Indels	Gaps
QY	1 RTIATKPNFVSTTISKSPFAPPRIOPSFGENVGKAGALLFSVNLTPVENVGQVYYPV 60	100.0%; Score 699; DB 1; Length 154;	154	0	0	0	0
DB	19 RTIATKPNFVSTTISKSPFAPPRIOPSFGENVGKAGALLFSVNLTPVENVGQVYYPV 78						
QY	61 YBDDYGLGRVNTPADASOSITVIVDEKGMKMDGASVTPNQQTTFKLANNTSGEKKI 120						
DB	79 YBDDYGLGRVNTPADASOSITVIVDEKGMKMDGASVTPNQQTTFKLANNTSGEKKI 138						
QY	121 SPEIYNDQVWVGYYVN 136						
DB	139 SPEIYNDQVWVGYYVN 154						

```

ID      F6a1_ECOLI      STANDARD;      PRT;      154 AA.
AC      P53508;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-MAY-2005 (Rel. 47, Last annotation update)
DE      CS6 fimbrial subunit A precursor (CS6 pilin).
DN
GN      Name=csaA;
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562;
RN
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=O167:HS / E10703 / EIEC;
RA      Wolf M.K., de Haan L., Casseels F.C., Willshaw G.A., van Gestel E.,
RA      Gaesstra W., Warren R., Boedeker E.C.;
RL      Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
CC      -1- FUNCTION: Fimbriae (also called pili), polar filaments radiating
CC      from the surface of the bacterium to a length of 0.5-1.5
CC      micrometers and numbering 100-300 per cell, enable bacteria to
CC      colonize the epithelium of specific host organs.
CC
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC
DR      EMBL, U04844; AAC45093.1; -, Unassigned_DNA.
DR      PIR, I60266; I60266.
DR      KW      Fimbria, Signal.
FT      SIGNAL      1      18      Potential.
FT      CHAIN      19      154      CS6 fimbrial subunit A.
PT      SEQUENCE      154 AA; 16953 MW; B4801BE6B6BD206 CRC64;
SO
5. ry Match      93.7%; Score 685; DB 1; Length 154;
Beat Local Similarity 91.9%; Pred. No. 1.4e-52;
Matches 125; Conservative 8; Mismatches 3; Indels 0; Gaps 0.
Qy      1      RTEIATKQFPVSTTTSKSPFAPEPRIQSFGENNGEGALLFSVNLTVPENVSQVTVPV 60
Db      19      RTEIATKQFPVSTTTSKSPFAPEPRIQSFGENNGEGALLFSVNLTVPENVSQVTVPV 78
Qy      61      YDEYDGLRLVNTADASQSIYQIYDEYDGKKMLKDHGAELVTPNQOITTKALNTYSGEKI 120
Db      79      YDEYDGLRLVNTADDSQSIYQIYDDDKRKKLKDHGAELVTPNQOITTKALNTYSGDKXI 138
Qy      121      SPGIYNDQVMVGYGVN 136
Db      139      PPGIYNDQVMVGYGVN 154
RESULT 3
F6a1_ECOLI
ID      F6a1_ECOLI      STANDARD;      PRT;      167 AA.
AC      P53510;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-MAY-2005 (Rel. 47, Last annotation update)
DE      CS6 fimbrial subunit B precursor.
DN
GN      Name=csaB;
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562;
RN
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=O167:HS / E10703 / EIEC;
RA      Wolf M.K., de Haan L., Casseels F.C., Willshaw G.A., van Gestel E.,
RA      Gaesstra W., Warren R., Boedeker E.C.;
RL      Submitted (JAN-1994) to the EMBL/Genbank/DBJ databases.
CC
CC
CC

```

```

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U04844; AAC45094.1; -; Unassigned_DNA.
DR PIR; I83348; I83348.
KW Fimbria; Signal.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 167 CS6 fimbrial subunit B.
SQ SEQUENCE 167 AA; 18022 MW; F9FPIB9IC411848E CRC64;
Query Match 17.2%; Score 120.5; DB 1; Length 167;
Best Local Similarity 30.1%; Pred. No. 0.0047;
Matches 43; Conservative 24; Mismatches 61; Indels 15; Gaps 7;
QY 7 KNFPEVSTTIKSPFAPE---PRIQPSFGENVKEGALLFSVNLTVPENVSQTVYP-- 59
Dy 27 KSLDVVNNIENQNTL-PIDSAVALIIPNVYSDPKLSQLYTVENTIIPAGSAVKIAPTDS 85
QY 60 VYDEYDGLGLVNTADASQSIYQIYDKE--GKKMLKDHA-EVTPNOQTTFALANTSG 116
Dy 86 LTSGGQOIGLTVANNPDQNNNTYIRDSGAGRFMAGQKGSFVKENTSYTFSAL-YTGG 144
QY 117 E---KKISPGIYNDQVWNGYYVN 136
Dy 145 EYPSNGYSYSGTYAGHLTVSPYSN 167
RESULT 4
F6B2_ECOLI STANDARD; PRT; 167 AA.
ID F6B2_ECOLI
AC P53511;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE CS6 fimbrial subunit B precursor.
DN Name=cs6B;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NC NCBI_TaxId=562;
RX [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=B8775;
RC Wolf M.K., de Haan L.A.M., Cassels F.C., Willschaw G.A., Geesbel E.C.M.,
RA Gaastera W., Warren R., Boedeker E.C.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U04846; AAB51362.1; -; Unassigned_DNA.
DR Fimbria; Signal.
KW Fimbria; Signal.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 167 CS6 fimbrial subunit B.
SQ SEQUENCE 167 AA; 17930 MW; B8DD041D3BBFBDGB CRC64;
Query Match 16.2%; Score 113.5; DB 1; Length 167;
Best Local Similarity 28.5%; Pred. No. 0.021;
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;
QY 7 KNFPEVSTTIKSPFAPE---PRIQPSFGENVKEGALLFSVNLTVPENVSQTVYP-- 59
Dy 27 KSLDVVNNIENQNTL-PIDSAVALIIPNVYSDPKLSQLYTVENTIIPAGSAVKIAPTDS 85
QY 60 VYDEYDGLGLVNTADASQSIYQIYDKEGKKMLKDHA-----EVTPNOQTTF 108
Dy 86 LTSGGQOIGLTVANNPDQNNNTYIRDSGAGRFMAGQKGSFVKENTSYTFSAL-YTGG 137

```

QY 109 KALNTYSGE---KKISPGIYNDQVGVYV 136  
DB 138 SAI-YTGEYFNSGSGTYAGNLTVSFYN 167  
RESULT 5  
ID 054U88\_DICDI PRELIMINARY; PRT; 440 AA.  
AC 054U88;  
DT 13-SEP-2005 (TEMBLrel. 31, Created)  
DT 13-SEP-2005 (TEMBLrel. 31, Last sequence update)  
DE 13-SEP-2005 (TEMBLrel. 31, Last annotation update)  
DE Hypothetical protein.  
ORFNames=DD80215209;  
OS Dicyostelium discoidium (Slime mold).  
OC Eukaryota; Mycetozoa; Dicyostelida; Dicyostelium.  
NCBI\_TaxID=44689;  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=AX4;  
RA Eichinger L., Pachbat J.A., Gloeckner G., Rajandream M.-A.,  
RA Tunggal B., Bertman M., Song J., Olsen R., Szafranski K., Xu Q.,  
RA Bankler A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,  
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bacon N.,  
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,  
RA Cooper J., Haydock S., van Drieseche N., Cronin A., Goodhead I.,  
RA Muzny D., Moutier T., Pain A., Lu M., Harper D., Lindsay R.,  
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchteser C.,  
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,  
RA Louieged H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,  
RA Urushihara H., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
RA Ma J., Kohara Y., Harper S., Simmonds M., Spiegler S., Tivey A.,  
RA Sugano S., White B., Walker D., Woodward J., Winkler T., Tanaka Y.,  
RA Shatsky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,  
RA Chaitin R.L., Glabe R., Loomis W.F., Plutzer M., Kay R.R.,  
RA Williams J., Dear P.H., Noegel A.A., Barrett B., Kuspa A.,  
RT "The genome of the social amoeba Dicyostelium discoidium";  
RL Nature 0:0-0(2005).  
CC -1- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC EMBL; AAFI0100071; EAL66967.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 440 AA; 50234 MW; FC85D88231E1A694 CRC64;  
Query Match 13.2%; Score 92; DB 2; Length 440;  
Best Local Similarity 30.3%; Pred. No. 6.4;  
Matches 36; Conservative 16; Mismatches 33; Indels 34; Gaps 7;  
QY 44 VNLTVENSGVTVYVYDDEYGLRVNTADASQ-----SIYQ-----IYDEK 88  
DB 259 LNTLPLEISQETVF-----YWEPKLNTIDYQFNPNVIFDNNSDVYISIPQIDVK 312  
QY 89 GKXGLKDHGALEVPN---OOITFKALNTYS-----GEKISPGIYNDQ---VMGVYV 136  
DB 313 GVATIGTHSGNVLENMSKTLFKKENVYDKVGVGE--YMPGLNDKEGHSVYCHTN 369  
RESULT 6  
ID 072P30\_LEPIC PRELIMINARY; PRT; 918 AA.  
AC 072P30;  
DT 05-JUL-2004 (TEMBLrel. 27, Created)  
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)  
DE Membrane carboxypeptidase.  
ORFNames=LI12646;  
OS Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar  
OS Copenhagen).  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.

OX NCBI\_TaxID=44275;  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=Flocruz LI-130;  
RC PubMed=15028702; DOI=10.1128/JB.186.7.2164-2172.2004;  
RX Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,  
RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartshorn R.A.,  
RA Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H.,  
RA Coutinho L.L., Degraive W.M., Dellagostin O.A., El-Dorri H.,  
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gambertini M., Gigliotti E.A.,  
RA Gees-Neco A., Goldman G.H., Goldman M.H.S., Harakava R.,  
RA Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,  
RA Kuranai E.E., Lemos E.G.M., Lemos M.V.P., Marino C.L., Nunes L.R.,  
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schirfer A.,  
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.U.G., Ferro J.A.,  
RA Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Slyke M.A.,  
RT "Comparative genomics of two Leptospira interrogans serovars reveals  
RT novel insights into physiology and pathogenesis";  
RL J. Bacteriol. 186:2164-2172(2004).  
DR EMBL; AEO17297; AAS71206.1; -; Genomic\_DNA.  
DR GO; GO:0009274; C:cell wall (sensu Bacteria); IEA.  
DR GO; GO:0004180; F:carboxypeptidase activity; IEA.  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0006558; F:penicillin binding; IEA.  
DR GO; GO:0009252; F:peptidoglycan biosynthesis; IEA.  
DR InterPro; IPR001264; Glyco\_trans\_51.  
DR InterPro; IPR001460; Pencl\_bind\_tpept.  
DR Pfam; PF00912; Transglyl; 1.  
DR Pfam; PF00905; Transpeptidase; 1.  
DR ProDom; PD001895; Glyco\_trans\_51; 1.  
KW Carboxypeptidase; Complete proteome.  
SQ SEQUENCE 918 AA; 102807 MW; 2D9E521B0396960 CRC64;  
Query Match 12.4%; Score 87; DB 2; Length 918;  
Best Local Similarity 25.9%; Pred. No. 44;  
Matches 37; Conservative 25; Mismatches 57; Indels 24; Gaps 6;  
QY 6 TKNPVSTTISKSPFAPPEPPIQSPFSENGKSGALLFSVNLTVENSGVTVYVYDDEY 65  
DB 478 TGFSPYTYTVEBPVQALPPIVAVYNDVQKNG-LVRKRLTIDNNKSSSEYVRRYQDL 536  
QY 66 -GLRLVNT-----ADASQSIYQIYDEKGMKLDHGALEVPNQIIFKA---LNTY 114  
DB 537 SPALPELPIDPSGGQNESGLQVLAVPDSTGELIMHGS-----EFKADNQDLRT 589  
QY 115 SGEKK-----ISPGIYNDQVVG 132  
DB 590 TANKROTGSIKRILYSALIEG 612  
RESULT 7  
ID 08F7D7\_LEPIN PRELIMINARY; PRT; 918 AA.  
AC 08F7D7;  
DT 01-MAR-2003 (TEMBLrel. 23, Created)  
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)  
DE Penicillin-binding protein 1A (EC 3.4.-.-) (EC 2.4.2.-).  
GN OrderedLocuNames=LA1009;  
OS Leptospira interrogans.  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
OX NCBI\_TaxID=173;  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
RX MEDLINE=22598143; PubMed=12712204; DOI=10.1038/nature01597;  
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,  
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Yu Y.-F.,  
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,  
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,  
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,  
RA Saint Girons I., Somerville R.L., Wen Y.-M., Shi M.-H., Chen Z.,  
RA Xu J.-G., Zhao G.-P.;



"Unique physiological and pathogenic features of *Leptospira* interorgans revealed by whole-genome sequencing.";  
 RT Nature 422:888-893(2003).  
 DR EMBL; AE011285; AAM48208.1; -; Genomic DNA.  
 DR GO; GO:0009274; C:cell wall (sensu Bacteria); IEA.  
 DR GO; GO:0003824; F:catalytic activity; IEA.  
 DR GO; GO:0008658; F:penicillin binding; IEA.  
 DR GO; GO:0009252; F:peptidoglycan biosynthesis; IEA.  
 DR InterPro; IPR001264; Glyco\_trans\_51.  
 DR InterPro; IPR001460; Penic\_bind\_type.  
 DR Pfam; PF00912; Transgly\_1.  
 DR Pfam; PF00905; Transpeptidase; 1.  
 DR ProDom; PD001895; Glyco\_trans\_51; 1.  
 DR Complete proteome.  
 SO SEQUENCE 918 AA; 102818 MW; 5EEA6FD48DC58326 CRC64;  
 Query Match 12.4%; Score 87; DB 2; Length 918;  
 Best Local Similarity 25.9%; Pred. No. 44;  
 Matches 37; Conservative 25; Mismatches 57; Indels 24; Gaps 6;  
 QY 6 TKNFEVSTTISKSPAPRPRIQPSFGENVGKGLLFVNLTPENVSQVTVPYDEDDY 65  
 DB 478 TCGFSTYTTTSPVOAELPKIVKNVDVQKNG-LVRRKRLTDNKNSETAVFRRIIDL 536  
 QY 66 --GLGLRVNT-----ADASQSIIVQIVDEKGRKMLKHGAETVNTQUTFKA--LNYT 114  
 DB 537 SPALFLFTDTSFGQNSGGLVALVADPSTGILLMHGGS-----EPRADNQLDRT 589  
 QY 115 SGEKK-----ISPGIYNDQVWVG 132  
 DB 590 TAMKQTGSSIKPIIYSAIETG 612  
 RESULT 8  
 ID 081IA3\_BACCR PRELIMINARY; PRT: 749 AA.  
 AC 081IA3\_BACCR PRELIMINARY;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 GN Formate acetyltransferase (EC 2.3.1.54).  
 GN OrderedLocusNames=BC0491;  
 GN Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
 OC Bacillus cereus group.  
 NC NCB1\_TaxID=226900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;  
 RA Ivanova N., Sorokin A., Anderson I., Gallen N., Candelson B.,  
 RA Kaputir V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Gotsman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Pusch G., Haselkorn R., Fontein M., Ehrlich S.D.,  
 RA Overbeek R., Kyriides N.C.;  
 "Genome sequence of *Bacillus cereus* and comparative analysis with  
*Bacillus anthracis*.";  
 RL Nature 423:87-91(2003).  
 DR EMBL; AE016999; AAP07529.1; -; Genomic DNA.  
 DR HSSP; P09373; 1H16.  
 DR GO; GO:0005737; C:cyttoplasm; IEA.  
 DR GO; GO:0008415; F:acetyltransferase activity; IEA.  
 DR GO; GO:0008861; F:formate C-acetyltransferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0005975; F:carbohydrate metabolism; IEA.  
 DR GO; GO:0006006; P:glucose metabolism; IEA.  
 DR InterPro; IPR005949; Form\_actrans.  
 DR InterPro; IPR001150; Form\_actrans.GR.  
 DR InterPro; IPR004184; Pyr\_Form\_lyase.  
 DR Pfam; PF01228; Gly\_radical; 1.  
 DR Pfam; PF02901; PFL; 1.  
 DR TIGRFAMs; TIGR01255; Pyr\_form\_ly\_1; 1.  
 DR PROSITE; PS00850; GLY\_RADICAL; 1.

KW Acyltransferase; Complete proteome; Transferase.  
 SO SEQUENCE 749 AA; 84681 MW; 52E5AE51D0661002 CRC64;  
 Query Match 11.9%; Score 83; DB 2; Length 749;  
 Best Local Similarity 22.4%; Pred. No. 81;  
 Matches 35; Conservative 28; Mismatches 67; Indels 26; Gaps 5;  
 QY 4 IATON-FPVSTTISKSPAPRPRIQPSFGENVG--KGLLFVNLTPENVSQVTVYP 59  
 DB 344 LVTKNSRFLTLTDLNGLPAPRPNLTLMVSKQLPENFKRYCARMSIKTSALQYENDDIRP 403  
 QY 60 VYDEDDYGLRLVNT-----ADASQSIIVQIVDEKGRKMLKHGAETVNTQOI 106  
 DB 404 EYGDYGIACCVSAMRGKMQPFGAPANLAKALLVINGKDKSKRQVGPETPA---I 460  
 QY 107 TFKALNTSGEKKIS-----PGIYNDQVWVYVN 136  
 DB 461 TSEVLNTEVVMHKFDMTWMELAGLYLNTLVNTHYMH 496  
 RESULT 9  
 ID 08MO08\_CAEEL PRELIMINARY; PRT: 12268 AA.  
 AC 08MO08\_CAEEL PRELIMINARY;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Hypothetical protein K07E12.1.  
 GN ORFNames=K07E12.1;  
 GN Caenorhabditis elegans.  
 OC Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=Bristol NZ;  
 RC MEDLINE=99069613; PubMed=9851916;  
 RX The C. elegans sequencing consortium;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; U00054; AAM48546.1; -; Genomic DNA.  
 DR HSSP; P01130; 1H28.  
 DR Ensembl; K07E12.1; Caenorhabditis elegans.  
 DR WormBase; WBGene0019500; K07E12.1.  
 DR WormPeP; K07E12.1b; CE31032.  
 DR InterPro; IPR006209; EGF\_1like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003598; IG\_c2.  
 DR InterPro; IPR008162; Pyrophosphatase.  
 DR InterPro; IPR000436; Suhl\_SCR\_CCP.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00008; EGF; 1.  
 DR Pfam; PF00041; fn3; 10.  
 DR Pfam; PF00047; Ig; 3.  
 DR Pfam; PF00084; Suhl; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR SMART; SM00032; CCP; 1.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00060; FN3; 10.  
 DR SMART; SM00408; IGF2; 1.  
 DR SMART; SM00327; VWA; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS50853; FN3; 10.  
 DR PROSITE; PS50835; IG\_1like; 7.  
 DR PROSITE; PS00387; PRASE; UNKNOWN\_1.  
 DR PROSITE; PS50234; VWF; 2.  
 KW Complete proteome; Hypothetical protein.  
 SO SEQUENCE 12268 AA; 1282637 MW; 4418C6C048E635A6 CRC64;  
 Query Match 11.9%; Score 83; DB 2; Length 12268;

```

DR PROSITE; PS01166; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PSS0853; FN3; 11.
DR PROSITE; PSS0835; IG_LIKE; 7.
DR PROSITE; PS00387; EPASE; UNKNOWN_1.
DR PROSITE; PSS0234; VMPA; 4.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 13100 AA; 1373302 MW; E13D0CBF28C7EAEFE CRC64;

Query Match 11.9%; Score 83; DB 2; Length 13100;
Best Local Similarity 25.6%; Pred. No. 27;
Matches 34; Conservative 23; Mismatches 50; Indels 28; Gaps 6;

Qy 10. PVESTTISKSPFAFBPR--QPSFGENVKGGALLFSVNLVPEVNSQVTVYPPVYDEBYGL 67
Db 10446 PLGDSSTGYTTEGQLVGDEEGKFPVGPQVL-----PTDSAGHYVPI---TGA 10499
Qy 68 GRLVNTFADAQSIIYQIVLDEKGGKMLKD-----HGAETVFNQITFFKALNTSGEKI 120
Db 10495 DRQILITDAGKRPITVFNEDGIQLPTDSSGYALGHGELVPESTNGVPLN-----KDG 1054

Qy 121 SPGIYNDQVNVY 135
Db 10550 TPLPTNDS---GHFV 10561

RESULT 11
Q6H0E5_BACAN PRELIMINARY; PRT; 248 AA.
AC 06H0E5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Methionine aminopeptidase (MAP) (Peptidase M).
OS OrderedLocuNames=BAS5204;
GN Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group.
OX NCBI_TaxID=1392;
RN [1]
RP NCLECTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Sterne;
RA Brectin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus anthracis Sterne.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
EMBL: AB017325; AAT57493.1; -; Genomic DNA.
DR GO:0004339; F:methionyl aminopeptidase activity; IEA.
DR GO:0008233; F:peptidase activity; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR001714; Pept_M24_MAP1.
DR InterPro: IPR002467; Pept_M24_MAP1.
DR InterPro: IPR000994; Peptidase_M24.
DR Pfam: PF00557; Peptidase_M24; 1.
DR PRINTS; PR00599; MAPEPTIDASE.
DR TIGRFAMs; TIGR00500; met_pdae.1; 1.
KW Aminopeptidase; Cobalt; Hydrolase; Protease.
SQ SEQUENCE 248 AA; 27062 MW; 14A6A817CE3BF9DE CRC64;

Query Match 11.7%; Score 82; DB 2; Length 248;
Best Local Similarity 25.6%; Pred. No. 27;
Matches 23; Conservative 18; Mismatches 29; Indels 20; Gaps 2;

Qy 67 IGRLVNTA-----DASQSIIYQIVDEKGGKMLKDHAETVFNQITFFKALNTSGEKI 120
Db 15 IGRIVALLAREMKKAKPGKTTKELDLIGKVLDERGALSAPEKEDYFPVGTCSVNEEV 74

Qy 121 SPGIYNDQVNVY-----GYNV 136
Db 75 AHGIPGDRVLKEGDVLNVVDVSAALDGYAD 104

```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: April 10, 2006, 10:13:10 ; Search time 189 Seconds  
(without alignment)  
316.167 Million cell updates/sec

Title: US-10-768-093-9  
Sequence: 1 RTIATKMPVSTTISKFF.....EKKISPGYNDVWVGYVN 136

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues  
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_21.\*  
1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*  
9: geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	699	100.0	136	ADRI4916
2	699	100.0	136	ADRI4916
3	699	100.0	154	ADRI4912
4	699	100.0	154	ADRI4912
5	113.5	16.2	146	ADRI4917
6	113.5	16.2	146	ADRI4917
7	113.5	16.2	167	ADRI4913
8	84.5	12.1	1239	ADRI4913
9	81.5	11.7	610	ADRI4913
10	81.5	11.7	610	ADRI4913
11	79.5	11.4	835	ADRI4913
12	78.5	11.2	332	ADRI4913
13	78.5	11.2	332	ADRI4913
14	78.5	11.2	335	ADRI4913
15	78.5	11.2	335	ADRI4913
16	78.5	11.2	335	ADRI4913
17	78.5	11.2	335	ADRI4913
18	77.5	11.1	754	ADRI4913
19	77.5	11.1	356	ADRI4913
20	77.5	11.1	356	ADRI4913
21	76.5	10.9	186	ADRI4913
22	76.5	10.9	186	ADRI4913
23	76.5	10.9	186	ADRI4913
24	76.5	10.9	186	ADRI4913

25	75.5	10.8	246	ABU39391
26	75.5	10.8	358	ABU39391
27	75.5	10.8	351	ABU39391
28	75.5	10.8	603	ABU39391
29	75.5	10.7	130	ADH87135
30	75.5	10.7	640	ABU39391
31	75.5	10.7	736	ABU39391
32	75.5	10.7	736	ABU39391
33	74.5	10.7	400	AA682840
34	74.5	10.7	497	ADSO6735
35	74.5	10.7	497	ADSO6735
36	74.5	10.7	750	ADSO6735
37	73.5	10.5	357	ABU39054
38	73.5	10.5	556	ABU39054
39	73.5	10.5	580	ADN46850
40	73.5	10.4	736	AAU35880
41	72.5	10.4	375	ADA89505
42	72.5	10.4	645	ADY23668
43	72.5	10.4	1773	ABU3908
44	72.5	10.4	2541	ABU3908
45	72.5	10.3	218	ADA35599

ALIGNMENTS

RESULT 1  
ID ADRI4916 standard; protein; 136 AA.  
AC ADRI4916;  
DT 04-NOV-2004 (first entry)  
DE CS6 related protein #1.  
KW CS6; cs6A; cs6B; cs6C; cs6D; origin of replication; lac promoter;  
KM kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli.  
XX Unidentified.  
OS US2004156829-A1.  
PN 12-AUG-2004.  
PD 12-AUG-2004.  
PF 02-FEB-2004; 2004US-00768093.  
PR 13-MAY-1994; 94US-00243482.  
PR 24-JAN-1997; 97US-00788145.  
PR 10-JAN-2000; 2000US-00479877.  
(US9A ) US SEC OF ARMY.  
PI Wolf MK, Casella FU, Boedeker EC;  
XX WPI; 2004-592722/57.  
DR WPI; 2004-592722/57.  
PT Inducing the production of antibodies against CS6 protein, useful for  
PT preventing pathological effects of enterotoxigenic E. coli, by  
PT administering a composition comprising transformed bacteria producing CS6  
PT antigens.  
PS Claim 4; SEQ ID NO 9; 22pp; English.  
XX The invention relates to a method of inducing, in a susceptible host, the  
XX production of antibodies against a CS6 protein, comprising administering  
XX a composition of matter comprising bacteria transformed with a plasmid  
XX containing genes cs6A, cs6B, cs6C and cs6D, an origin of replication,  
XX lac promoter and a kanamycin resistance gene, where the bacteria  
XX expresses both CS6A and CS6B proteins. The invention also relates to a  
XX composition of matter comprising a protein in a pharmaceutical carrier.  
XX The pharmaceutical carrier is a carbonated liquid. The method is useful  
XX for inducing the production of antibodies against the CS6 protein. The

protein and composition are used as vaccines for preventing pathological effects of enterotoxigenic E. coli. This sequence represents a CS6 related protein of the invention.

Sequence 136 AA;

Query Match 100.0%; Score 699; DB 8; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.1e-72;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RTEATKMPVSTTSSKSPFAPEPRIOPSFGENVGKGLLFSVNLTPENVSOVTYYPV 60  
DB 1 RTEATKMPVSTTSSKSPFAPEPRIOPSFGENVGKGLLFSVNLTPENVSOVTYYPV 60  
OY 61 YDEBYGLRLVNTADASOSIIYQIVDEKGMKLDHGAETVPNOQITFKALNTSGEKKI 120  
DB 61 YDEBYGLRLVNTADASOSIIYQIVDEKGMKLDHGAETVPNOQITFKALNTSGEKKI 120  
OY 121 SPGIYNDQVMVGYVN 136  
DB 121 SPGIYNDQVMVGYVN 136

RESULT 2

ADM47737 standard; protein; 136 AA.

ADM47737;

21-APR-2005 (first entry)

E. coli CS6 csa mature protein.

antibacterial; vaccine; DNA purification; colonization factor antigen.

Escherichia coli.

US2005025787-A1.

03-FEB-2005.

12-JAN-2004; 2004US-00754641.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(WOLF/) WOLF M K.

(CASS/) CASSELS F J.

(BOED/) BOEDEKER E C.

Wolf MK, Cassele FJ, Boedecker EC;

WPI; 2005-131784/14.

Inducing in a susceptible host, the production of antibodies against CS6

protein, useful for stimulating protective antibodies against Escherichia

coli by administering a composition comprising bacteria transformed with

a plasmid.

Claim 4; SEQ ID NO 9; 23pp; English.

The invention relates to a method of inducing in a susceptible host, the production of antibodies against CS6 protein comprising giving a composition of matter made of bacteria transformed with a plasmid, which contain genes csaA and csaB, all of csaC and DNA sequence csaD that encodes at least 802 amino acids (at least 2406 base pairs), an origin of replication, a lac promoter, and a kanamycin resistance gene, where the bacteria overexpress both CS6A and CS6B proteins. The method and proteins are useful for stimulating protective antibodies against enterotoxigenic Escherichia coli. This sequence corresponds to the E. coli CS6 csaA mature protein

Sequence 136 AA;

Query Match 100.0%; Score 699; DB 9; Length 136;  
Best Local Similarity 100.0%; Pred. No. 1.1e-72;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RTEATKMPVSTTSSKSPFAPEPRIOPSFGENVGKGLLFSVNLTPENVSOVTYYPV 60  
DB 1 RTEATKMPVSTTSSKSPFAPEPRIOPSFGENVGKGLLFSVNLTPENVSOVTYYPV 60  
OY 61 YDEBYGLRLVNTADASOSIIYQIVDEKGMKLDHGAETVPNOQITFKALNTSGEKKI 120  
DB 61 YDEBYGLRLVNTADASOSIIYQIVDEKGMKLDHGAETVPNOQITFKALNTSGEKKI 120  
OY 121 SPGIYNDQVMVGYVN 136  
DB 121 SPGIYNDQVMVGYVN 136

RESULT 3

ADRI4912 standard; protein; 154 AA.

ADRI4912;

04-NOV-2004 (first entry)

Escherichia coli E8775 antigen csaA protein.

CS6; csaA; csaB; csaC; csaD; origin of replication; lac promoter; kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli;

Escherichia coli E8775.

Escherichia coli.

US2004156829-A1.

12-AUG-2004.

02-FEB-2004; 2004US-00768093.

13-MAY-1994; 94US-00243482.

24-JAN-1997; 97US-00788145.

10-JAN-2000; 2000US-00479877.

(USSA ) US SEC OF ARMY.

Wolf MK, Cassele FJ, Boedecker EC;

WPI; 2004-592722/57.

N-PSDB; ADRI4911.

Inducing the production of antibodies against CS6 protein, useful for preventing pathological effects of enterotoxigenic E. coli, by administering a composition comprising transformed bacteria producing CS6 antigens.

Disclosure; SEQ ID NO 5; 22pp; English.

The invention relates to a method of inducing, in a susceptible host, the production of antibodies against a CS6 protein, comprising administering a composition of matter comprising bacteria transformed with a plasmid containing genes csaA, csaB, csaC and csaD, an origin of replication, a lac promoter and a kanamycin resistance gene, where the bacteria express both CS6A and CS6B proteins. The invention also relates to a composition of matter comprising a protein in a pharmaceutical carrier. The pharmaceutical carrier is a carbonated liquid. The method is useful for inducing the production of antibodies against the CS6 protein. The protein and composition are used as vaccines for preventing pathological effects of enterotoxigenic E. coli. This sequence represents the Escherichia coli E8775 antigen csaA of the invention.

Sequence 154 AA;

Query Match 100.0%; Score 699; DB 8; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-72;  
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC 1 RTEIATKNFPVSTTISKSFPAPRIQPSFGENVKSGALLFSVNLTPENVSGVTVYPV 60  
 XX |||||||  
 SQ RTEIATKNFPVSTTISKSFPAPRIQPSFGENVKSGALLFSVNLTPENVSGVTVYPV 78  
 Db 19 RTEIATKNFPVSTTISKSFPAPRIQPSFGENVKSGALLFSVNLTPENVSGVTVYPV 78  
 QY 61 YDEBYGRLVNTADASQSIITVQVDEKGGKMLKDGAETVPNOQITFKALNTSGEKI 120  
 Db 79 YDEBYGRLVNTADASQSIITVQVDEKGGKMLKDGAETVPNOQITFKALNTSGEKI 138  
 QY 121 SPGIYNDQVMVGYVYN 136  
 Db 139 SPGIYNDQVMVGYVYN 154

RESULT 4  
 ADM47733  
 ID ADM47733 standard; protein; 154 AA.

AC ADM47733;  
 DT 21-APR-2005 (first entry)  
 XX  
 DE E. coli CS6 csaA full length protein.  
 KM antibacterial; vaccine; DNA purification; colonization factor antigen.  
 OS Escherichia coli.

Key Location/Qualifiers  
 FT Peptide 1..18  
 FT /note= "signal peptide"  
 FT Protein 19..154  
 FT /note= "mature peptide"

US2005025787-A1.

PN 03-FEB-2005.  
 PF 12-JAN-2004; 2004US-00754641.  
 PR 13-MAY-1994; 94US-00243482.  
 PR 24-JAN-1997; 97US-00788145.  
 PR 10-JAN-2000; 2000US-00479877.

XX (WOLF/) WOLF M K.  
 PA (CASS/) CASSELS F J.  
 XX (BOED/) BOEDEKER E C.

PI Wolf MK, Casseels FJ, Boedecker EC;  
 XX  
 DR WPI; 2005-131784/14.  
 DR N-PSDB; ADM47732.

PT Inducing in a susceptible host, the production of antibodies against CS6  
 PT protein, useful for stimulating protective antibodies against Escherichia  
 PT coli by administering a composition comprising bacteria transformed with  
 PT a plasmid.  
 XX

PS Disclosure; SEQ ID NO 5; 23pp; English.

CC The invention relates to a method of inducing in a susceptible host, the  
 CC production of antibodies against CS6 protein comprising giving a  
 CC composition of matter made of bacteria transformed with a plasmid, which  
 CC contain genes csaA and csaB, all of csaC and DNA sequence csaD that  
 CC encodes at least 802 amino acids (at least 2406 base pairs), an origin of  
 CC replication, a lac promoter, and a kanamycin resistance gene, where the  
 CC bacteria overexpress both CS6A and CS6B proteins. The method and proteins  
 CC are useful for stimulating protective antibodies against enterotoxigenic  
 CC Escherichia coli. This sequence corresponds to the E. coli CS6 csaA

CC protein  
 XX  
 SQ Sequence 154 AA;

Query Match 100.0%; Score 699; DB 9; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-72;  
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNFPVSTTISKSFPAPRIQPSFGENVKSGALLFSVNLTPENVSGVTVYPV 60  
 Db 19 RTEIATKNFPVSTTISKSFPAPRIQPSFGENVKSGALLFSVNLTPENVSGVTVYPV 78  
 QY 61 YDEBYGRLVNTADASQSIITVQVDEKGGKMLKDGAETVPNOQITFKALNTSGEKI 120  
 Db 79 YDEBYGRLVNTADASQSIITVQVDEKGGKMLKDGAETVPNOQITFKALNTSGEKI 138  
 QY 121 SPGIYNDQVMVGYVYN 136  
 Db 139 SPGIYNDQVMVGYVYN 154

RESULT 5

ADR14917  
 ID ADR14917 standard; protein; 146 AA.

AC ADR14917;  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE CS6 related protein #2.

KM CS6; csaA; csaB; csaC; csaD; origin of replication; lac promoter;  
 KM kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli.

OS Unidentified.

US2004156829-A1.

12-AUG-2004.

PF 02-FEB-2004; 2004US-00768093.  
 PR 13-MAY-1994; 94US-00243482.  
 PR 24-JAN-1997; 97US-00788145.  
 PR 10-JAN-2000; 2000US-00479877.

PA (USSA ) US SEC OF ARMY.

PI Wolf MK, Casseels FJ, Boedecker EC;  
 XX  
 DR WPI; 2004-592722/57.

PT Inducing the production of antibodies against CS6 protein, useful for  
 PT preventing pathological effects of enterotoxigenic E. coli, by  
 PT administering a composition comprising bacteria transformed bacteria producing CS6  
 PT antigens.  
 XX

PS Claim 6; SEQ ID NO 10; 22pp; English.

CC The invention relates to a method of inducing, in a susceptible host, the  
 CC production of antibodies against a CS6 protein, comprising administering  
 CC a composition of matter comprising bacteria transformed with a plasmid  
 CC containing genes csaA, csaB, csaC and csaD, an origin of replication, a  
 CC lac promoter and a kanamycin resistance gene, where the bacteria  
 CC express both CS6A and CS6B proteins. The invention also relates to a  
 CC composition of matter comprising a protein in a pharmaceutical carrier.  
 CC The pharmaceutical carrier is a carbonated liquid. The method is useful  
 CC for inducing the production of antibodies against the CS6 protein. The  
 CC protein and composition are used as vaccines for preventing pathological  
 CC effects of enterotoxigenic E. coli. This sequence represents a CS6  
 CC related protein of the invention.

XX Sequence 146 AA;  
 SQ

Query Match 16.2%; Score 113.5; DB 8; Length 146;  
Best Local Similarity 28.5%; Pred. No. 0.00014;  
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

OY 7 KNFVSTTISKFFAPE-----PRIQPSFGENVGKGGALLFSVNLVTPENVQVTVYP--- 59  
DB 6 KSLDVNNVNIENFI-PDIDSAVRIIPVNYDSPKLDLSQLYTEVMTIPAGVSAVKIAPIDS 64  
OY 60 VYDEBYGRLVNTADASQSIITYQIVDEKGMKLDHGA-----EYTPNQITF 108  
DB 65 LTSSGQOIGKLVNVPDQNNYYI-----RKDSGAGNFMAGQKGFVKENTSYTF 116  
OY 109 KALNTSGE---KKISPGIYNDQVWVGYVN 136  
DB 117 SAI-VTGGEPNPGYSSTGTAGNLTVSFYSN 146

RESULT 6  
ID ADR14913 standard; protein; 167 AA.  
XX ADR14913;  
AC ADR14913;  
DT 04-NOV-2004 (first entry)  
XX  
XX Escherichia coli E8775 antigen cs6B protein.  
DE  
XX  
XX CS6; cs6A, cs6B, cs6C, cs6D, origin of replication; lac promoter;  
KW kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli;  
KW Escherichia coli E8775.  
XX  
XX Escherichia coli.  
OS  
XX US2004156829-A1.  
XX  
XX 12-AUG-2004.  
PD  
XX PF 02-FEB-2004; 2004US-00768093.  
PR  
XX PR 13-MAY-1994; 94US-00243482.  
PR 24-JAN-1997; 97US-00788145.  
PR 10-JAN-2000; 2000US-00479877.  
XX  
XX (USSA ) US SEC OF ARMY.  
PA  
XX PI Wolf MK, Casseels FU, Boedeker EC;  
XX  
XX WPI; 2004-592722/57.  
DR  
XX PT Inducing the production of antibodies against CS6 protein, useful for  
PT preventing pathological effects of enterotoxigenic E. coli, by  
PT administering a composition comprising transformed bacteria producing CS6  
PT antigens.  
PS  
XX Disclosure; SEQ ID NO 6; 22pp; English.  
XX  
XX The invention relates to a method of inducing, in a susceptible host, the  
XX production of antibodies against a CS6 protein, comprising administering  
XX a composition of matter comprising bacteria transformed with a plasmid  
XX containing genes cs6A, cs6B, cs6C, cs6D, an origin of replication, a  
XX lac promoter and a kanamycin resistance gene, where the bacteria  
XX express both CS6A and CS6B proteins. The invention also relates to a  
XX composition of matter comprising a protein in a pharmaceutical carrier.  
XX The pharmaceutical carrier is a carbonated liquid. The method is useful  
XX for inducing the production of antibodies against the CS6 protein. The  
XX protein and composition are used as vaccines for preventing pathological  
XX effects of enterotoxigenic E. coli. This sequence represents the  
XX Escherichia coli E8775 antigen cs6B of the invention.  
XX  
XX Sequence 167 AA;  
SQ

Query Match 16.2%; Score 113.5; DB 8; Length 167;

Best Local Similarity 28.5%; Pred. No. 0.00017;  
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

OY 7 KNFVSTTISKFFAPE-----PRIQPSFGENVGKGGALLFSVNLVTPENVQVTVYP--- 59  
DB 27 KSLDVNNVNIENFI-PDIDSAVRIIPVNYDSPKLDLSQLYTEVMTIPAGVSAVKIAPIDS 85  
OY 60 VYDEBYGRLVNTADASQSIITYQIVDEKGMKLDHGA-----EYTPNQITF 108  
DB 86 LTSSGQOIGKLVNVPDQNNYYI-----RKDSGAGNFMAGQKGFVKENTSYTF 137  
OY 109 KALNTSGE---KKISPGIYNDQVWVGYVN 136  
DB 138 SAI-VTGGEPNPGYSSTGTAGNLTVSFYSN 167

RESULT 7  
ID ADM47734 standard; protein; 167 AA.  
XX ADM47734;  
AC ADM47734;  
DT 21-APR-2005 (first entry)  
XX  
XX E. coli CS6 cs6B full length protein.  
DE  
XX  
XX antibacterial; vaccine; DNA purification; colonization factor antigen.  
KW  
XX  
XX Escherichia coli.  
OS  
XX US2005025787-A1.  
PN  
XX PD 03-FEB-2005.  
XX  
XX 12-JAN-2004; 2004US-00754641.  
PF  
XX PR 13-MAY-1994; 94US-00243482.  
PR 24-JAN-1997; 97US-00788145.  
PR 10-JAN-2000; 2000US-00479877.  
XX  
XX (WOLF/) WOLF M K.  
PA (CASS/) CASSELS F J.  
PA (BOED/) BOEDEKER E C.  
XX  
XX Wolf MK, Casseels FU, Boedeker EC;  
XX  
XX WPI; 2005-131784/14.  
DR  
XX PT Inducing in a susceptible host, the production of antibodies against CS6  
PT protein, useful for stimulating protective antibodies against Escherichia  
PT coli by administering a composition comprising bacteria transformed with  
PT a plasmid.  
PS  
XX Disclosure; SEQ ID NO 6; 23pp; English.  
XX  
XX The invention relates to a method of inducing in a susceptible host, the  
XX production of antibodies against CS6 protein comprising giving a  
XX composition of matter made of bacteria transformed with a plasmid, which  
XX contain genes cs6A and cs6B, all of cs6C and DNA sequence cs6D that  
XX encodes at least 802 amino acids (at least 2406 base pairs), an origin of  
XX replication, a lac promoter, and a kanamycin resistance gene, where the  
XX bacteria overexpress both CS6A and CS6B proteins. The method and proteins  
XX are useful for stimulating protective antibodies against enterotoxigenic  
XX Escherichia coli. This sequence corresponds to the E. coli CS6 cs6B  
XX protein  
XX  
XX Sequence 167 AA;  
SQ

Query Match 16.2%; Score 113.5; DB 9; Length 167;  
Best Local Similarity 28.5%; Pred. No. 0.00017;  
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

OY 7 KNFVSTTISKFFAPE-----PRIQPSFGENVGKGGALLFSVNLVTPENVQVTVYP--- 59



Db 27 KSLDQVNVNIEQNF1-PDIDSAVRRIIPVNYDSDPKLDQSLYVEMTIRPAGVSAVXIAPTDS 85  
Qy 60 VYDEYDYGRLVNTADASQSI-YQIVDEKGMKLDHGA-----EVTNQGITF 108  
Db 86 LTSSGQOIGKLVNVPDQNNYI-----RKDSAGNFMAQKGSFVKENTSYTF 137  
Qy 109 KALNVTSGE---KKISPGIYNDQVWVGYVN 136  
Db 138 SAI-YTGGEYPNSGYSGTYAGNLTVSFYSN 167

RESULT 8  
ADR86415  
ID ADR86415 standard; protein; 1239 AA.  
XX  
AC ADR86415;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Aspergillus fumigatus essential gene protein #465.  
XX  
KM Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;  
XX drug screening.  
XX  
OS Aspergillus fumigatus.  
XX  
PN W02004067709-A2.  
XX  
PD 12-AUG-2004.  
XX  
PF 16-JAN-2004; 2004WO-US001099.  
XX  
PR 17-JAN-2003; 2003US-0441281P.  
XX  
PR 13-JUN-2003; 2003US-0478196P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX (ELIT-) ELITRA PHARM INC.  
XX  
PI Jiang B, Hu W, Lemieux S, Roemer T;  
XX  
DR WPI; 2004-594200/57.  
XX  
DR N-PSDB; ADR85828.  
XX  
PT New purified or isolated Aspergillus fumigatus nucleic acid molecule  
XX encoding a gene product, useful for diagnosing and/or treating invasive  
XX fungal infections, such as Farmer's lung disease.  
XX  
PS Claim 1; SEQ ID NO 3465; 164bp; English.  
XX  
XX The present invention relates to Aspergillus fumigatus genes that are  
XX essential and are potential targets for drug screening. The methods and  
XX compositions of the present invention are useful for diagnosing and/or  
XX treating invasive Aspergillus fumigatus infection. Including the allergic  
XX forms of the disease, such as Farmer's lung disease. They can also be  
XX used in various drug discovery purposes, such as expression of the  
XX recombinant protein, hybridization assay and construction of nucleic acid  
XX arrays. The present sequence represents an Aspergillus fumigatus  
XX essential gene protein sequence, used during diagnosis and drug  
XX development in the invention. These genes share a high degree of sequence  
XX conservation with known essential genes of candida albicans. The sequence  
XX data for this patent is not represented in the printed specification, but  
XX was obtained in electronic format from WIPO.

SO Sequence 1239 AA;  
Query Match 12.1%; Score 84.5; DB 8; Length 1239;  
Best Local Similarity 28.0%; Pred. No. 8;  
Matches 35; Conservative 20; Mismatches 53; Indels 17; Gaps 6;

Qy 12 STTISKFPAP-----EPRIQPSFGENVGKGLLFSVNLVTPENVQVTVYVYDE 63  
Db 1031 ATTIVNVQSPAAENGAKTAVSEP---EGTAKERSARTLGLMNVPTVDARIRAMV-E 1086

Qy 64 DYGLGELVNTADASQSI-YQIVDEKGMKLDHGA-EVTNQOI---TFKALNVTSGEK 118  
Db 1087 PYQPLIKIVLRPHQGIYVFAVDVNHAGKASLELEGGEIAPGQRLHVGTVSELLKQSAEK 1146  
Qy 119 KISPG 123  
Db 1147 KGGPG 1151

RESULT 9  
AEB37679  
ID AEB37679 standard; protein; 610 AA.  
XX  
AC AEB37679;  
XX  
DT 08-SEP-2005 (first entry)  
XX  
DE L. pneumophila protein SEQ ID NO 2011.  
XX  
KM detection; infection; Antibacterial; Vaccine.  
XX  
OS Legionella pneumophila.  
XX  
PN W02005049642-A2.  
XX  
PD 02-JUN-2005.  
XX  
PF 23-SEP-2004; 2004WO-IB003578.  
XX  
PR 21-NOV-2003; 2003FR-00013687.  
XX  
PA (INSP ) INST PASTEUR.  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX (UYLY-) UNIV LYON 1 BERNARD CLAUDE.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;  
XX Rusniok C, Bouchlier C, Zidane N, Magnier A, Kunet F, Vandenesch F;  
XX Jarraud S;  
XX  
DR WPI; 2005-388305/40.  
XX  
PT New genome of Legionella pneumophila Paris strain and derived  
XX polypeptides, useful for detection or identification of the strain and  
XX for treatment and prevention of infections.  
XX  
PS Claim 3; SEQ ID NO 2011; 660bp; English.  
XX  
XX The invention relates to an isolated or purified nucleotide sequences (I)  
XX from Legionella pneumophila Paris strain. (I), and their related  
XX sequences or fragments, are useful as primers and probes for detection  
XX and amplification, including differentiation between the Paris and  
XX Philadelphia strains of Legionella pneumophila and to prepare recombinant  
XX (hybrid) polypeptides (II). (II) are also useful for preparation of  
XX specific antibodies (Ab), also used for detection/identification of  
XX Legionella, and some (I), specifically those involved in synthesis of  
XX surface proteins, are targets for identification of inhibitors. (II), or  
XX vectors that contain (I), are useful as vaccines and immunogenic  
XX compositions, for treatment and prevention of infections by L.  
XX pneumophila. The present sequence represents the amino acid sequence of a  
XX L. pneumophila protein.

SO Sequence 610 AA;  
Query Match 11.7%; Score 81.5; DB 9; Length 610;  
Best Local Similarity 25.7%; Pred. No. 6.3; Indels 31; Gaps 5;  
Matches 29; Conservative 19; Mismatches 34;  
Qy 9 FPVST-----TISKSPAPRIQPSFGENVGKGLLFSVNLVTPENV 53  
Db 312 FPIVSDDFEAFREALKLSTNDASLFEPESSALRGFCFLGMLHMEI---IQRL- 367

OY 54 QVTVPYVDEYDGLGRLVNTADASQSIYQIVDEKXKMLKDGAETVPMQOI 106  
DB 368 -----EREYNLD-LISTA---PTVVYQIVTQGETLLIDNPSHLPTPTQI 408

RESULT 10  
ID AEB40990 standard; protein; 610 AA.  
AC AEB40990;  
XX  
XX 08-SEP-2005 (first entry)  
XX  
XX L. pneumophila protein SEQ ID NO 5322.  
XX  
XX detection; infection; Antibacterial; Vaccine.  
XX  
XX Legionella pneumophila.  
XX  
XX WO2005049642-A2.  
XX  
XX 02-JUN-2005.  
XX  
XX 23-SEP-2004; 2004WO-1B003578.  
XX  
XX 21-NOV-2003; 2003FR-00013687.  
XX  
XX (INSP ) INST PASTEUR.  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX (UTLY-) UNIV LYON 1 BERNARD CLAUDE.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Buchrieser C, Tichit M, Etienne J, Ma L, Cazale C, Glaser P,  
PI Ruenick C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F,  
PI Jarraud S;  
XX  
XX WPI; 2005-388305/40.  
XX  
XX New genome of Legionella pneumophila Paris strain and derived  
PT polypeptides, useful for detection or identification of the strain and  
PT for treatment and prevention of infections.  
XX  
XX Claim 3; SEQ ID NO 5322; 660bp; English.

XX The invention relates to an isolated or purified nucleotide sequences (I)  
XX from Legionella pneumophila Paris strain. (I), and their related  
XX sequences or fragments, are useful as primers and probes for detection  
XX and amplification, including differentiation between the Paris and  
XX Philadelphia strains of Legionella pneumophila and to prepare recombinant  
XX (hybrid) polypeptides (II). (II) are also useful for preparation of  
XX specific antibodies (Ab), also used for detection/identification of  
XX Legionella, and some (I), specifically those involved in synthesis of  
XX surface proteins, are targets for identification of inhibitors. (II), or  
XX vectors that contain (I), are useful as vaccines and immunogenic  
XX compositions, for treatment and prevention of infections by L.  
XX pneumophila. The present sequence represents the amino acid sequence of a  
XX L. pneumophila protein.  
XX  
XX Sequence 610 AA:  
SQ

Query Match 11.7%; Score 81.5; DB 9; Length 610;  
Best Local Similarity 25.7%; Pred. No. 6.3;  
Matches 29; Conservative 19; Mismatches 34; Indels 31; Gaps 5;

OY 9 FVPSV-----TISKSPFAPRPPROPSFGENVKKGALLFSVVLTVPMVMS 53  
DB 312 FVPSDDDEAPFEALAKSLANDASLFPBPESSEALGFGRCGFLMLMEI---IQEHL- 367

OY 54 QVTVPYVDEYDGLGRLVNTADASQSIYQIVDEKXKMLKDGAETVPMQOI 106  
DB 368 -----EREYNLD-LISTA---PTVVYQIVTQGETLLIDNPSHLPTPTQI 408

RESULT 11  
ID ABP26319 standard; protein; 835 AA.  
AC ABP26319;  
XX  
XX 02-JUL-2002 (first entry)  
XX  
XX Streptococcus polypeptide SEQ ID NO 1814.  
XX  
XX Streptococcus pyogenes; GAS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
XX Streptococcus agalactiae.  
XX  
XX WO200234771-A2.  
XX  
XX 02-MAY-2002.  
XX  
XX 29-OCT-2001; 2001WO-GB004789.  
XX  
XX 27-OCT-2000; 2000GB-00026333.  
XX 24-NOV-2000; 2000GB-00028727.  
XX 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit y Rosi, Grandi G, Frazer C,  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
XX N-PSDB; ABN66950.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
XX Claim 1; Page 3331; 4525bp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
XX streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
XX the specification. The proteins have antibacterial and antiinflammatory  
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX antibodies that bind (I) are used in the manufacture of medicaments for  
XX the treatment or prevention of infection or disease caused by  
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX biological sample. (I) is used to determine whether a compound binds to  
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX used as a vaccine or diagnostic composition. The disease caused by  
XX Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX acid encoding (I) may be used to recombinantly produce (I) and may be  
XX used in gene therapy. Antibodies to (I) are used for affinity  
XX chromatography, immunoassays, and distinguishing/identifying  
XX Streptococcus proteins  
XX  
XX Sequence 835 AA:  
SQ

Query Match 11.4%; Score 79.5; DB 5; Length 835;  
Best Local Similarity 21.8%; Pred. No. 17;  
Matches 29; Conservative 28; Mismatches 31; Indels 45; Gaps 7;

OY 13 TTISKSPFAPRPPROPSFGENVKKGALLFSVNLTVPMVMS-QVTVPYVDEYDGLGRLV 71  
DB 412 TPTLKTTPAPKPKP-----HVSISEKINYSVHPV-----LV 446

OY 72 NTADASQSIYQIVDEKXKMLKDGAETVPMQOITPKA-----LNTSGEKISFGI 124  
DB 447 PAAKPSKAVI-----DEAGQSV---NGKTVLPNALNVAKODFQYQKMTSOKIANKF 499

OY 125 -----YNDQWVG 132  
Db 500 VFIDYKDDALDG 512

RESULT 12  
AAG10781  
ID AAG10781 standard; protein; 332 AA.  
AC AAG10781,  
XX  
XX 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9239.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridization assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131444P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 14-MAY-1999; 99US-0134768P.  
PR 18-MAY-1999; 99US-0134941P.  
PR 19-MAY-1999; 99US-0135124P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137503P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 16-JUN-1999; 99US-0139455P.  
PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 26-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140981P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148314P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.

PR 16-AUG-1999; 99US-0149366P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 29-OCT-1999; 99US-0162142P.

## Query Match

Best Local Similarity 27.3%; Score 78.5; DB 3; Length 332;  
Matches 33; Conservative 20; Mismatches 39; Indels 29; Gaps 6;

DB 10 PVSTTI--SKSFAPDPRIQPSFGENVGKALLFESVNTLPBNSQVTVFVYDDEY-- 65  
DB 193 FVGSVIVSSKFRITAPRWLRKTLGGGMRQIGVLCALVALHENVAKL-----EDDHKK 246  
DB 66 -----GIGRL-----VNTDAQSITTYOVDEKGR-----KMKDHGAETVPNQITF 108  
DB 247 ARVLAEGNIRLRNVNAVETNITTYDIPEDPKFGAEACKSLDVGVLVIP--QATF 304

XX 109 K 109  
DB 305 R 305  
RESULT 13  
AA010780  
ID AA010780 standard; protein; 355 AA.  
XX  
AC AA010780;  
XX  
DT 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9238.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 25-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147193P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.

PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158223P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160747P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 25-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 26-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161982P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

## Query March

Best Local Similarity 27.3%; Score 78.5; DB 3; Length 355;  
Matches 33; Conservatave 20; Mismatches 39; Indels 29; Gaps 6;

Qy 10 PVSTTI--SKSFAPBRIQPSFGENVKSGALLFSVNLTVPENVSQVTVPTVDEY-- 65  
DB 216 PVGSVIVGSKKFIKRWLRKTLGGGWRQIGVLCMAALVALHENVAKL-----EDHKX 269  
Qy 66 -----GLGRH-----VNTADASOSIITVYDEKK-----KVLKHGAETVNTQVTF 108

DB 270 ARVLAEGLNRIERLRVNAVAETNIITVYDIPEDPKFGAEACKSLDVGVLVIP--QATF 327  
 OY 109 K 109  
 DB 328 R 328

RESULT 14  
 AEA36463  
 ID AEA36463 strand; protein; 355 AA.  
 XX  
 AC AEA36463;  
 XX  
 DT 28-JUL-2005 (first entry)  
 XX  
 XX A. thaliana chromosome 3 threonine aldolase.  
 DE  
 KM threonine aldolase; silencing; mitochondrial targeting sequence;  
 KM transgenic; soybean; alternative oxidase; potato; formate dehydrogenase;  
 KM Nicotiana glauca; mitochondrial ATP synthase beta-subunit;  
 KM animal feed; foodstuff.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO2005047472-A2.  
 XX  
 PD 26-MAY-2005.  
 XX  
 PF 12-NOV-2004; 2004WO-US037369.  
 XX  
 PR 10-NOV-2003; 2003US-0519313P.  
 PA (BOYC-) BOYCE THOMPSON INST PLANT RES INC.  
 PI Jander G, Joshi V;  
 XX WPI; 2005-372364/38.  
 DR N-PSDB; AEA36464.  
 XX  
 PT New nucleic acid construct comprising a nucleic acid molecule,  
 PT mitochondrial targeting, promoter, and terminator sequences, useful for  
 PT increasing threonine content in plant seeds.  
 XX  
 PS Claim 2; SEQ ID NO 3; 130bp; English.

This sequence represents a threonine aldolase. This threonine aldolase is an example of a threonine aldolase which may be silenced by the nucleic acid construct of the invention which comprises a nucleic acid molecule configured to silence threonine aldolase expression or encodes a threonine aldolase, a mitochondrial targeting sequence, a 5' DNA promoter sequence, and a 3' terminator sequence, where the nucleic acid molecule, the mitochondrial targeting sequence, the promoter, and the terminator are operatively coupled to permit transcription of the nucleic acid molecule. This nucleic acid construct may be used in the generation of transgenic plants which have increased levels of seed threonine and/or isoleucine. The mitochondrial targeting sequence comprises a mitochondrial targeting leader sequence, e.g., a soybean alternative oxidase gene, a potato formate dehydrogenase gene, or a Nicotiana glauca mitochondrial ATP synthase beta-subunit gene. The nucleic acid molecule may comprise a dominant negative mutation and encode a nonfunctional threonine aldolase, resulting in suppression or interference of endogenous mRNA encoding threonine aldolase. The nucleic acid molecule is positioned in the nucleic acid construct to result in suppression or interference of endogenous mRNA encoding threonine aldolase. The nucleic acid molecule may encode threonine aldolase and be in the same orientation, or the nucleic acid molecule is an antisense form or a threonine aldolase encoding nucleic acid molecule. The mutation may be caused by a chemical mutagenizing agent, e.g., ethylmethanesulphonate, radiation, e.g., ultraviolet rays, gamma rays, or fast neutrons, or by inserting an inactivating nucleic acid molecule into the gene encoding the functional threonine aldolase or its promoter under conditions to inactivate the gene, where the deactivating nucleic acid molecule is a transposable element, e.g., an Activator (Ac) transposon, a

CC Dissociator (Ds) Transposon, or a Mutator (Mu) transposon. Increasing  
 CC isoleucine content in a plant comprises providing a non-wild-type plant  
 CC having increased threonine content compared to a corresponding wild-type  
 CC plant, transforming the non-wild-type plant with a nucleic acid construct  
 CC described above under conditions to yield a transgenic plant that over  
 CC expresses an enzyme that functions to catalyze biosynthesis of isoleucine  
 CC from threonine, and growing the transgenic plant under conditions to  
 CC increase the isoleucine content of the transgenic plant. The nucleic acid  
 CC construct and methods are useful for increasing threonine content in  
 CC plants, particularly in plant seeds, which are then used in producing  
 CC animal feed and foodstuff.  
 XX  
 SQ Sequence 355 AA;  
 Query Match 11.2%; Score 78.5; DB 9; Length 355;  
 Best Local Similarity 27.3%; Pred. No. 6.3;  
 Matches 33; Conservative 20; Mismatches 39; Indels 29; Gaps 6;

OY 10 PUSTT--SKSFPAEPRIQSPFGENVKKGALLFSVLTVPENVSQTVVPVDEDY-- 65  
 DB 216 PVGSVTVGSKKFTYKARWLRTLLGGKMKQIGVLCNALVALHENVAKL-----EDDHKK 269  
 OY 66 -----GKRL-----VNTADASQSIYQIVDEKSK-----KMLKDGAETVPPNQITTF 108  
 DB 270 ARVLAEGLNRIERLRVNAVAETNIITVYDIPEDPKFGAEACKSLDVGVLVIP--QATF 327  
 OY 109 K 109  
 DB 328 R 328

RESULT 15  
 AAG10779  
 ID AAG10779 standard; protein; 367 AA.  
 XX  
 AC AAG10779;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 9237.  
 XX  
 KM Protein identification; signal transduction pathway; metabolic pathway;  
 KM hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 PF 25-FEB-2000; 2000EP-00301439.  
 XX  
 XX 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127462P.  
 PR 06-APR-1999; 99US-0128234P.  
 PR 08-APR-1999; 99US-0128714P.  
 PR 16-APR-1999; 99US-0128845P.  
 PR 19-APR-1999; 99US-0130077P.  
 PR 21-APR-1999; 99US-0130449P.  
 PR 23-APR-1999; 99US-0130510P.  
 PR 23-APR-1999; 99US-0130891P.  
 PR 28-APR-1999; 99US-0131448P.  
 PR 30-APR-1999; 99US-0132048P.  
 PR 30-APR-1999; 99US-0132407P.  
 PR 04-MAY-1999; 99US-0132484P.  
 PR 05-MAY-1999; 99US-0132485P.  
 PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134222P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136762P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138004P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 06-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143547P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147322P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152353P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156586P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158359P.  
PR 13-OCT-1999; 99US-0159283P.  
PR 13-OCT-1999; 99US-0159284P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159846P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2006, 10:17:05 ; Search time 39 Seconds  
(without alignments)  
335.525 Million cell updates/sec

Title: US-10-768-093-9

Sequence: 1 RTETATKNEFVSTTISKSF.....EKISPGIYNDQWVGYYVN 136

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR 80: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	655	93.7	154 2 160266	CS6 structural sub
2	120.5	17.2	167 2 183348	CS6 structural sub
3	83	11.9	13055 2 T16580	hypothetical prote
4	81	11.6	595 2 AF2097	hypothetical prote
5	78	11.2	608 2 AB1293	probable flagellar
6	77.5	11.1	740 1 JC6010	formate C-acetyltr
7	77.5	11.1	749 2 AB9785	hypothetical prote
8	77.5	11.1	4936 2 AH2515	hypothetical prote
9	77	11.0	356 2 T25354	hypothetical prote
10	76.5	10.9	569 2 T13531	probable porphyrin
11	76	10.9	557 2 F69481	probable acid-CoA
12	75.5	10.8	361 2 T49882	pectin methyl-este
13	75	10.7	266 2 AB2851	chaperone protein
14	75	10.7	736 2 DE4534	DNA topoisomerase
15	75	10.7	1472 2 H82802	fibriillar assembly
16	74.5	10.7	946 2 UCS575	inner-alpha-trypt
17	74	10.6	275 2 JC7604	CS66 spliced varia
18	73.5	10.5	521 2 SE2794	probable lipoprote
19	73.5	10.5	556 2 AI0134	probable phage pro
20	73.5	10.5	718 2 T71930	hypothetical prote
21	73.5	10.5	881 2 F69438	conserved hypotet
22	73	10.4	736 2 B11972	DNA topoisomerase
23	73	10.4	1487 2 AG2560	hypothetical prote
24	73	10.4	1843 2 SI1803	collagen alpha 1(V
25	72.5	10.4	925 2 S50480	hypothetical prote
26	72	10.3	880 2 AF2128	hypothetical prote
27	71.5	10.2	357 2 S71772	naringenin 3-dioxy
28	71.5	10.2	357 2 S61415	naringenin 3-dioxy
29	71.5	10.2	744 2 F75625	serine proteinase,

30	71.5	10.2	812 2 T01618	hypothetical prote
31	71	10.2	318 2 T50039	beta-1,4-galactosy
32	71	10.2	1157 2 F97255	fusion of alpha-gl
33	70.5	10.1	217 2 T29388	hypothetical prote
34	70.5	10.1	250 2 F72247	methionine aminope
35	70.5	10.1	323 2 S69647	hypothetical prote
36	70.5	10.1	357 1 G64053	choistamate synthas
37	70.5	10.1	430 2 AG1381	enolase homolog en
38	70.5	10.1	430 2 AH1750	thiophen and furan
39	70.5	10.1	458 2 F84157	dihydropterolate sy
40	70.5	10.1	498 2 C90413	conserved membrane
41	70.5	10.1	896 2 AB1156	probable lipoprote
42	70.5	10.1	1644 2 AC0823	toxin B - Clostrid
43	70.5	10.1	2366 2 SI0317	hypothetical prote
44	70.5	10.1	2911 2 T20566	outer membrane pro
45	70	10.0	242 2 H82720	

ALIGNMENTS

RESULT 1  
160266  
CS6 structural subunit A - Escherichia coli  
C/Species: Escherichia coli  
C/Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 09-Jul-2004  
C/Accession: 160266  
R/Willshaw, G.A.; Smith, H.R.; McConnell, M.M.; Rowe, B.  
FEMS Microbiol. Lett. 49, 473-478, 1988  
A/Title: Cloning of genes encoding coli-surface (CS) antigens in enterotoxigenic Escheri-  
A/Reference number: 160266  
A/Accession: 160266  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-154 <RES>  
A/Cross-references: UNIPROT:P53508; UNIPARC:UP1000012A400; EMBL:U04844; NID:9442375; PID

Query Match  
Best Local Similarity 93.7% Score 655, DB 2; Length 154;  
Best Local Similarity 91.7% Pred. No. 1.4e-54;  
Matches 125; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 RTETATKNEFVSTTISKSFAPPRLOPSYGENYKGEKGLLFYNLWPNVPSOVTVYVPV 60  
DB 19 RTETATKNEFVSTTISKSFAPPRLOPSYGENYKGEKGLLFVSLVY PENVSVTVYVPV 78  
QY 61 YDEPDYGLRLVNTADQSIITIVDEKGRKMDHGAETVPOQITFKALNTYSGEKI 120  
DB 79 YDEPDYGLRLVNTADQSIITIVDEKGRKMDHGAETVPOQITFKALNTYSGEKI 138  
QY 121 SPGIYNDQWVGYYVN 136  
DB 138 SPGIYNDQWVGYYVN 154

RESULT 2  
183348  
CS6 structural subunit B - Escherichia coli  
C/Species: Escherichia coli  
C/Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 09-Jul-2004  
C/Accession: 183348  
R/Willshaw, G.A.; Smith, H.R.; McConnell, M.M.; Rowe, B.  
FEMS Microbiol. Lett. 49, 473-478, 1988  
A/Title: Cloning of genes encoding coli-surface (CS) antigens in enterotoxigenic Escheri-  
A/Reference number: 160266  
A/Accession: 183348  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-167 <RES>  
A/Cross-references: UNIPROT:P53510; UNIPARC:UP1000012A402; EMBL:U04844; NID:9442375; PID

Query Match  
Best Local Similarity 17.2% Score 120.5, DB 2; Length 167;  
Best Local Similarity 30.1% Pred. No. 0.00041;  
Matches 43; Conservative 24; Mismatches 61; Indels 15; Gaps 7;

[illegible]

C:Comment: This enzyme must be activated by the lysase-specific pyruvate formate-lyase-activating enzyme (Pfl) and causes peptide cleavage.

C:Gene: pfl

C:Start codon: TTG

C:Complex: homodimer

A:Description: catalyzes the reversible conversion of acetyl-CoA and formate into pyruvate and formate

A:Pathway: anaerobic glucose metabolism

C:Superfamily: formate C-acetyltransferase 1; glycyl radical homology

C:Keywords: acyltransferase; coenzyme A; homodimer; lysase; thiol-ester bond

F:681-740/Domain: glycyl radical homology <GPR>

F:405/Active site: Cys (cysteine thiol radical intermediate) #status predicted

F:406/Active site: Cys (S-acetylcysteine intermediate) #status predicted

F:715/Active site: Gly (stable glycyl radical) #status predicted

Query Match 11.1%; Score 77.5; DB 1; Length 740;

Best Local Similarity 25.0%; Pred. No. 31;

Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

4 IATKN-FPVSTTISKSPFAPRIPQSPFGENVKGALLFSVNLTPENVQV---TVY 58

336 LVTKNSRRLHTLNLGSAPEPMTVMSENL-BESFKKFAEMSLTDSIQYENDDIR 394

59 PVYDEYGL-----GRLVNTADASQSIYQIVDEKGGKMLKHGAETP 102

395 PIVDDVAIACVSAVRGKMOFGARCNLAKLLINSGVDK-----KGIVKP 447

103 N-QOITFKALNTSGEKKISPGIYND-QVMGVYV 136

448 DIEPITDEVLIDY---EKVKENYFVLEVMAGLYV 479

RESULT 7

Formate acetyltransferase [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C:Accession: E89785

R:Kuroda, M.; Ohya, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mutant, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirata, K.

C:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; PMID:21311952; PMID:11418146

A:Accession: E89785

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-749 <KUR>

A:Cross-references: UNIPROT:Q99WZ7; UNIPARC:UPI00000D769B; GB:BA000018; PID:g13700141; F

A:Experimental source: strain N315

C:Genetics:

A:Gene: pflB

C:Superfamily: formate C-acetyltransferase 1; glycyl radical homology

Query Match 11.1%; Score 77.5; DB 2; Length 749;

Best Local Similarity 23.8%; Pred. No. 31;

Matches 40; Conservative 24; Mismatches 45; Indels 59; Gaps 9;

4 IATKN-FPVSTTISKSPFAPRIPQSPFGENVKGALLFSVNLTPENV----- 52

344 LVTKNSRRLHTLNLGSAPEPMTVMSENL-BESFKKFAEMSLTDSIQYENDDIR 394

53 -----SQVTVVPVYDEYGLGRVNT-----ADASQSIYQIVDEKGGK 92

390 TSSIQYENDDIRRESYGGDYGIACVSAVMTTGKMOFGARCNLAKLLINSGVDK- 448

93 LKHGAETP-N-QOITFKALNTSGEKKISPGIYNDQVM---VGVYV 136

DB 449 ---SGAUGVRFEGINSEVLEVDVFKKF-----DQMDMLAGYV 467

RESULT 8

AH2515

hypothetical protein alr7304 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C:Accession: AH2515

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuch

Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; PMID:21595285; PMID:11759840

A:Accession: AH2515

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4936 <KUR>

A:Cross-references: UNIPROT:Q8YKJ3; UNIPARC:UPI000011021D; GB:BA000020; PIDN:BA078388.1;

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr7304

A:Genome: plasmid

Query Match 11.1%; Score 77.5; DB 2; Length 4936;

Best Local Similarity 25.0%; Pred. No. 3,4e+02;

Matches 42; Conservative 19; Mismatches 46; Indels 59; Gaps 8;

2 TEATKNFPVSTT-----ISKSPFAPRIPQSPFGENVKGALLFSVNL 47

2406 TGLIKDDPTTVDEGIGSAQWTFIPNLAFERATQYISG-----GLSYKENGTTVT 2459

48 VPENVSQVTVTP-----VYDEYGLGRVNTADASQSIYQ---IYDEKGGK 91

2460 VPLLSTPIYVPOAEIYDFHQRDVPADDPFTNDIET-----SVPSLAIVLRNEGGK 2514

92 MLKHGAETPNQOITFKALNTSGEKKI---SPGIYNDQVMGVYV 136

2515 EAKN-----LKITSQPKYDNEKGLLIDPQIISFVN 2547

RESULT 9

hypothetical protein T27D1.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T25354

R:Sims, M.

submitted to the EMBL Data Library, February 1995

A:Reference number: Z20021

A:Accession: T25354

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-356 <WIL>

A:Cross-references: UNIPROT:Q09638; UNIPARC:UPI000013BF59; EMBL:Z48245; PIDN:CAA86290.1;

A:Experimental source: clone T27D1

C:Genetics:

A:Gene: CESP:T27D1.3

A:Map position: 3

A:Insertion: 86/2; 118/2; 164/1; 233/1; 280/2; 322/2

C:Superfamily: vertebrate rhodopsin

Query Match 11.0%; Score 77; DB 2; Length 356;

Best Local Similarity 28.8%; Pred. No. 14;

Matches 19; Conservative 15; Mismatches 30; Indels 2; Gaps 1;

52 VSGVTVVPVYDEYGLGRVNTADASQSIYQIVDEKGGKMLKD--HGAETPNQOITFK 109

DB 282 VLVNTIHPYDPIILMAALMTGSAANPLVTLFSQFRRLRVLCPSDVENETKTY 341

QY 110 ALNTYS 115

DB 342 SINNTS 347

RESULT 10

T43531  
 probable potassium channel chain n2p2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
 C:Accession: T43531; T22711  
 R:Wang, Z.W.; Salkoff, L.  
 submitted to the EMBL Data Library, August 1998  
 A:Description: Potassium channels in C. elegans.  
 A:Reference number: 222450  
 A:Accession: T43531  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-569 <MAN>  
 A:Cross-references: UNIPROT:Q20817; UNIPARC:UPI000007FD46; EMBL:AF083654; PIDD:AAC32865.  
 R:Harria, B.  
 submitted to the EMBL Data Library, August 1996  
 A:Reference number: 219604  
 A:Accession: T22711  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-569 <ML>  
 A:Cross-references: UNIPARC:UPI000007FD46; EMBL:Z78198; PIDD:CAM01568.1; GSPDB:GN00023;  
 A:Experimental source: clone F55C5  
 C:Genetics:  
 A:Gene: F55C5.3  
 A:Map position: 5  
 A:Introns: 43/3; 140/1; 171/2; 294/2; 338/3; 382/2; 508/2  
 Query Match 10.9%; Score 76.5; DB 2; Length 569;  
 Best Local Similarity 30.8%; Pred. No. 27;  
 Matches 20; Conservative 12; Mismatches 28; Indels 5; Gaps 1;  
 Db 468 TWSTVDEDEGKRLQISQSLDRILFLVDEHKSMLEKRTQYPS-----KADKYQ 522  
 QY 116 GEKXI 120 .  
 Db 523 TQGRV 527  
 RESULT 11  
 F69481  
 probable acid-CoA ligase (EC 6.2.1.-) - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
 C:Accession: F69481  
 R:Klienx, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.;  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Meldrum, J.F.; McDonald, L.  
 Nature 390, 364-370; 1997  
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: F69481  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-557 <KLE>  
 A:Cross-references: UNIPROT:Q28423; UNIPARC:UPI0000056B85; GB:AE000975; GB:AE000782; NID  
 C:Superfamily: Probable acyl-CoA ligase medium chain; acetate-CoA ligase homology  
 C:Keywords: acid-thiol ligase  
 F:78-546/Domain: acetate-CoA ligase homology <ACLU>  
 Query Match 10.9%; Score 76; DB 2; Length 557;  
 Best Local Similarity 27.2%; Pred. No. 30;  
 Matches 31; Conservative 22; Mismatches 31; Indels 30; Gaps 7;  
 Db 295 VALKHPMSQSLFDLTKTFAQGSAPPPTWTFELKRNRIEPMNMGON---EGTGLFSY 351  
 QY 45 NLTVBNVSGVTTVYVDEYDGLRLVNTADASQSIITYIVDEKGMKLDHGA 98

Db 352 DRTIPDLKRAKSPPIPK-----VNDLPFPALRIKIYDESGKEM-KEPGS 397  
 RESULT 12  
 T9982  
 pectin methyl-esterase-like protein - Arabidopsis thaliana  
 N:Alternate names: protein T211.140  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
 A:Accession: T9982  
 R:Beyan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: 224493  
 A:Accession: T9982  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-361 <BEV>  
 A:Cross-references: UNIPROT:Q9LY17; UNIPARC:UPI0000048AB; EMBL:AL163912; GSPDB:GN00063;  
 A:Experimental source: cultivar Columbia; BAC clone T211  
 C:Genetics:  
 A:Gene: ATSP:T211.140  
 A:Map position: 5  
 A:Introns: 68/3; 161/3; 229/3; 309/2  
 Query Match 10.8%; Score 75.5; DB 2; Length 361;  
 Best Local Similarity 28.0%; Pred. No. 19;  
 Matches 28; Conservative 16; Mismatches 23; Indels 33; Gaps 6;  
 QY 48 VPENVSGV-----TVYVYDEYDGLRLVNTADASQSIITYIVDEKGMKLDHGA 101  
 Db 28 IPENRAQIPQWFKTNVKEYSQKGTLPALRAEAARQII--TVNQKG-----GA--- 75  
 QY 102 PNOQITFKALN-----YTSGEK-----KISPIYNDQVMV 131  
 Db 76 -----NFKTLBAKISFTGKNKRVIIYKLACGVNNEKVTI 110  
 RESULT 13  
 A82851  
 chaparrone protein precursor XF0082 (imported) - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
 C:Accession: A82851  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: For a complete list of authors see reference number A59328 below  
 A:Accession: A82851  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-266 <SIM>  
 A:Cross-references: UNIPROT:Q9PH64; UNIPARC:UPI00000C22C4; GB:AE003862; GB:AE003849; NID  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, E  
 de-Nero, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitching, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizy  
 Chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; de Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tanaka, M.H.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 A:Genetics:  
 A:Gene: XF0082  
 C:Superfamily: chaparrone protein papd

Query Match 10.7% Score 75; DB 2; Length 266;  
 Best Local Similarity 22.9%; Pred. No. 15;  
 Matches 41; Conservative 22; Mismatches 44; Indels 72; Gaps 9;

16 SXSFPAPF---RQPSFGENVGKGLLFSVNLTPPE-----NNSQVIVVPIYDD 64  
 78 SKAPFLVTPPIRVEPVGSQL---RLFFGTITLIPEDRESLFMLVLEIPPTPLEHSD 133  
 65 -----YGLGRLVNTADASQSI 81

134 KNAQTSADKSDSQNLQALFRSRIRKVFYRPATLKGANSPASLSKWNQNLVSNPT 193  
 82 -YQI-----VDEKGGKMLKDH-GAEVTPNQOITPKALNTYSG---EKISPEIYND 127  
 194 PYHVLTRTEALDDKGRKRTIDIKIGVLLTPGQETFEKA-NNDGLSSRKVSFSFYIND 251

RESULT 14  
 64534  
 DNA topoisomerase I - Helicobacter pylori (strain 26695)  
 C:Species: Helicobacter pylori  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #ext\_change 09-Jul-2004  
 C:Accession: D64534  
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A>Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; MUID:9794467; PMID:9252185  
 A:Accession: D64534  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-736 <TOM>  
 A:Cross-references: UNIPROT:P55991; UNIPARC:UPI000013713A; GB:AE00533; GB:AE00511; NID  
 C:Superfamily: DNA topoisomerase I

Query Match 10.7% Score 75; DB 2; Length 736;  
 Best Local Similarity 19.7%; Pred. No. 53;  
 Matches 23; Conservative 29; Mismatches 63; Indels 2; Gaps 1;

1 RTEITKMPVSTTSKSPAPPRIOPSFGENVGKGLLFSVNLTPPENVSQVTVIPV 60  
 228 KNELEKESYAISSIVKSKSPPT--PPMTSTLQOSASLLGFSPKTMSIAOKLYEGV 285  
 61 YDEDDGLRLVNTADASQSIYQIVDEKGGKMLKDHGAETPNQOITPKALNTSGE 117  
 286 ATPQGVMTVITMYRTDSLNIKALEARNKILKDYKDYLPKAKVYSSKNKMAOE 342

RESULT 15  
 H82802  
 Elmbtrial assembly protein XF0478 (Imported) - Xylella fastidiosa (strain 945C)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #ext\_change 09-Jul-2004  
 C:Accession: H82802  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82802  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1472 <SIM>  
 A:Cross-references: UNIPROT:Q9PG24; UNIPARC:UPI00000C2413; GB:AE003897; GB:AE003849; NID  
 A:Experimental source: strain 945C  
 R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H  
 as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fromm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvair  
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0478

Query Match 10.7% Score 75; DB 2; Length 1472;  
 Best Local Similarity 22.7%; Pred. No. 1.3e+02;  
 Matches 34; Conservative 22; Mismatches 40; Indels 54; Gaps 7;

11 VSTTSKSPAPPRIOPSF-----GENV-----GKEGALLFSVNLTPPENVS 53  
 790 VGASVSTSF-----VIRISSTHDGKDTGEMTRIRIDANGMLGDVLSASAGVPSGTS 845  
 54 QTVTVPYDEDYGLRLVNTADASQSIYQIVDEK-----GK 90  
 846 AIKKRVIT---ALSHVDITRA--SVVRRFVAELVSSSGDVATDAAGVGRFGYTPR 900  
 91 KMLKDHGAETPNQOITPKALNTSGEKKI 120  
 901 GVDDFGSSVTPNOLV-----NYLRGDKSM 925

Search completed: April 10, 2006, 10:21:26  
 Job time : 40 secs

! ;  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4742:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 amino acids  
TYPE: amino acid  
MOLECULE TYPE: linear  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8) LOCATION 1..378  
SEQUENCE DESCRIPTION: SEQ ID NO: 4742:  
US-09-107-532A-4742

Query Match 11.1%; Score 78.5; DB 2; Length 378;  
Best Local Similarity 27.1%; Pred. No. 0.56; Mismatches 48; Indels 27; Gaps 7;  
Matches 35; Conservative 19; Mismatches 48; Indels 27; Gaps 7;

13 TTISKSPFAPRIPQSPFGENGVK-EGALLFVNLTVEBNVSQVTVY-----TVY 61  
176 TSIHGRTPEAGAPAIK-VGKTEQVQEAKEVILPETIKVIOAMRSAPYEBAY 234  
62 D-----EDYGLRLVNTADASOSIYQIVDEKGMKLDHGAETV--PNOQITFK-- 109  
235 DLFALDEVEVMEGLGRV--GELPPEIFLFAFVEQVKEAFOLDGLRIYQPKXAKSSVKRI 291  
110 ALNVTSGEK 118  
292 AICGSGSEK 300

RESULT 3  
US-09-323-872A-25  
Sequence 25; Application US/09323872A  
Patent No. 6393539  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter  
TITLE OF INVENTION: Compositions and Methods for Bioremediation  
FILE REFERENCE: OHU-03640  
CURRENT APPLICATION NUMBER: US/09/323,872A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/072,433  
PRIOR FILING DATE: 1998-05-04  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 25  
LENGTH: 740

TYPE: PRT  
ORGANISM: Clostridium pasteurianum  
US-09-323-872A-25

Query Match 11.1%; Score 77.5; DB 2; Length 740;  
Best Local Similarity 25.0%; Pred. No. 2.1; Mismatches 60; Indels 35; Gaps 8;  
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

4 IATKN-PPVSTTIKSPFAPRIPQSPFGENGVK-EGALLFVNLTVEBNVSQV-----TVY 58  
336 LVTKNSFRYLTHTLNLGSAPEPNMTVLMSENL-DESFKKFCAEMSLTDSIQYENDDMR 394  
59 PVDYEDYGL-----GRLVNTADASOSIYQIVDEKGMKLDHGAETV 102  
395 PIVDDVYALACVSAEMRVKMDQPFGARCNLAKCLLAINGVDEK-----KGIKVVP 447  
103 N-QQITPFLNMTSGEKKISPGIYND-QVMVGYVN 136  
448 DIEPTDEVLDY---EKVENYFKVLEVMGLYVN 479

RESULT 4  
US-09-072-433-30  
Sequence 30; Application US/09072433  
Patent No. 6551814  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,433  
FILING DATE: 04-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: OHU-03344  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 740 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-072-433-30

Query Match 11.1%; Score 77.5; DB 2; Length 740;  
Best Local Similarity 25.0%; Pred. No. 2.1; Mismatches 60; Indels 35; Gaps 8;  
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

4 IATKN-PPVSTTIKSPFAPRIPQSPFGENGVK-EGALLFVNLTVEBNVSQV-----TVY 58  
336 LVTKNSFRYLTHTLNLGSAPEPNMTVLMSENL-DESFKKFCAEMSLTDSIQYENDDMR 394  
59 PVDYEDYGL-----GRLVNTADASOSIYQIVDEKGMKLDHGAETV 102  
395 PIVDDVYALACVSAEMRVKMDQPFGARCNLAKCLLAINGVDEK-----KGIKVVP 447









Db 886 RTEISTKEVPVQETKTTTYESFOIDGAG---GDSGLTLAQITT-SESVSTT 937

## RESULT 13

US-09-949-016-7359  
; Sequence 7359, Application US/09949016  
; Patent No. 6812319  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO010307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7359  
; LENGTH: 1008  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-7359

Query Match 10.3%; Score 72; DB 2; Length 1008;

Best Local Similarity 37.5%; Pred. No. 16; Mismatches 23; Indels 4; Gaps 2;  
Matches 21; Conservative 8;

Db 889 RTEISTKEVPVQETKTTTYESFOIDGAG---GDSGLTLAQITT-SESVSTT 940

## RESULT 14

US-09-134-000C-6740  
; Sequence 6740, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6740  
; LENGTH: 1467  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-6740

Query Match 10.2%; Score 71.5; DB 2; Length 1467;

Best Local Similarity 25.2%; Pred. No. 33; Mismatches 36; Indels 41; Gaps 7;  
Matches 33; Conservative 21;

Db 892 RTEISTKEVPVQETKTTTYESFOIDGAG---GDSGLTLAQITT-SESVSTT 937

Db 893 YGVEDTRLENFENIDEATFEDLTITTEFFKR-LRD-GFEYPDN-----NGET 939

Db 894 YGVEDTRLENFENIDEATFEDLTITTEFFKR-LRD-GFEYPDN-----NGET 939

Db 895 YGVEDTRLENFENIDEATFEDLTITTEFFKR-LRD-GFEYPDN-----NGET 939

RESULT 15  
US-09-538-092-213  
; Sequence 213, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Glaxo, Inc  
; APPLICANT: Mansfield, Tracy A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurataseqFormatter Version 0.9  
; SEQ ID NO 213  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number YDR480W  
US-09-538-092-213

Query Match 10.1%; Score 70.5; DB 2; Length 323;

Best Local Similarity 24.7%; Pred. No. 4.4; Mismatches 26; Indels 23; Gaps 5;  
Matches 23; Conservative 21;

Db 896 YGVEDTRLENFENIDEATFEDLTITTEFFKR-LRD-GFEYPDN-----NGET 939

Db 897 YGVEDTRLENFENIDEATFEDLTITTEFFKR-LRD-GFEYPDN-----NGET 939

Search completed: April 10, 2006, 10:22:19  
Job time : 47 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2006, 10:21:40 ; Search time 166 Seconds  
(without alignments)  
342.318 Million cell updates/sec

Title: US-10-768-093-9

Perfect score: 1 RTEIATKNFVSTTTSKSF.....EKISPGIYNDQWNGYYVN 136

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	699	100.0	136 4 US-10-768-093-9	Sequence 9, Appli
2	699	100.0	136 5 US-10-754-641-9	Sequence 9, Appli
3	699	100.0	154 4 US-10-768-093-5	Sequence 5, Appli
4	699	100.0	154 5 US-10-754-641-5	Sequence 5, Appli
5	113.5	16.2	146 4 US-10-768-093-10	Sequence 10, Appli
6	113.5	16.2	167 4 US-10-768-093-6	Sequence 6, Appli
7	113.5	16.2	167 5 US-10-754-641-6	Sequence 6, Appli
8	77.5	11.1	740 4 US-10-267-989-25	Sequence 25, Appli
9	77.5	11.1	740 4 US-10-357-567-25	Sequence 25, Appli
10	76	10.9	557 4 US-10-369-493-21384	Sequence 21384, A
11	76	10.9	2358 4 US-10-282-122A-68607	Sequence 68607, A
12	75.5	10.8	246 4 US-10-282-122A-67315	Sequence 67315, A
13	75	10.7	640 4 US-10-282-122A-44691	Sequence 44691, A
14	75	10.7	736 3 US-09-815-242-11293	Sequence 11293, A
15	75	10.7	736 4 US-10-282-122A-58624	Sequence 58624, A
16	74.5	10.7	249 4 US-10-424-599-250712	Sequence 250712, A
17	74.5	10.7	497 4 US-10-424-599-250712	Sequence 6030, Ap
18	74.5	10.7	750 5 US-10-501-282-3988	Sequence 3988, Ap
19	73.5	10.5	157 4 US-10-424-599-190153	Sequence 190153, A
20	73.5	10.5	357 4 US-10-282-122A-66978	Sequence 66978, A
21	73.5	10.5	356 4 US-10-282-122A-78182	Sequence 78182, A
22	73	10.4	302 4 US-10-335-977-7475	Sequence 7475, Ap
23	73	10.4	541 4 US-10-335-977-7476	Sequence 7476, Ap
24	72.5	10.4	736 3 US-09-815-242-11473	Sequence 11473, A
25	72.5	10.4	645 4 US-10-425-114-71452	Sequence 71452, A
26	72.5	10.4	1773 6 US-11-097-143-18516	Sequence 18516, A
27	72	10.3	274 4 US-10-369-493-8865	Sequence 8865, Ap

28	72	10.3	398 3 US-09-925-301-1264	Sequence 1264, Ap
29	72	10.3	398 4 US-10-264-049-2703	Sequence 2703, Ap
30	72	10.3	505 4 US-10-437-963-110397	Sequence 110397, A
31	72	10.3	520 4 US-10-369-493-11092	Sequence 11092, A
32	72	10.3	1005 5 US-10-921-707-1	Sequence 1, Appli
33	72	10.3	1705 4 US-10-437-963-138350	Sequence 138350, A
34	71.5	10.2	357 5 US-10-732-923-10066	Sequence 10066, A
35	71.5	10.2	357 5 US-10-732-923-10067	Sequence 10067, A
36	71.5	10.2	425 4 US-10-424-599-168174	Sequence 168174, A
37	71.5	10.2	571 4 US-10-282-122A-53248	Sequence 53248, A
38	71.5	10.2	657 4 US-10-282-122A-45312	Sequence 45312, A
39	71.5	10.2	1207 4 US-10-437-963-111041	Sequence 111041, A
40	70.5	10.1	357 4 US-09-815-242-10981	Sequence 10981, A
41	70.5	10.1	357 4 US-10-260-877-128	Sequence 128, App
42	70.5	10.1	357 4 US-10-268-611-41	Sequence 41, Appli
43	70.5	10.1	357 4 US-10-282-122A-58099	Sequence 58099, A
44	70.5	10.1	357 5 US-10-824-194-50	Sequence 50, Appli
45	70.5	10.1	430 4 US-10-282-122A-60954	Sequence 60954, A

ALIGNMENTS

RESULT 1  
US-10-768-093-9  
Sequence 9, Application US/10768093  
Publication No. US20040156829A1  
GENERAL INFORMATION:  
APPLICANT: Wolf, Marcia K  
Cassels, Fred J  
Boedeker, Edgar C  
TITLE OF INVENTION: Transformed Bacteria Producing Gs6  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Hendricks and Assoc  
STREET: P.O. Box 2509  
CITY: Fairfax  
STATE: VA  
COUNTRY: US  
ZIP: 22031  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
Application Number: US/10/768, 093  
Filing Date: 02-Feb-2004  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
Application Number: US/09/479, 877B  
Filing Date: 10-Jan-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna M  
REGISTRATION NUMBER: 32,535  
TELEPHONE: 703/425-8405  
TELEFAX: 703/425-8406  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 136 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-10-768-093-9  
Query Match 100.0%; Score 699; DB 4; Length 136;  
Best Local Similarity 100.0%; Pred. No. 5e-67;

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 60  
1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 60

QY 61 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 120  
61 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 120

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

## RESULT 2

US-10-754-641-9  
Sequence 9, Application US/10754641  
Publication No. US20050025787A1

GENERAL INFORMATION:

APPLICANT: U.S. Army Medical Research and Materiel Command

APPLICANT: Wolf, Marcia K.

APPLICANT: Boedeker, Frederick J.

TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES

FILE REFERENCE: 034047.033.3

CURRENT APPLICATION NUMBER: US/10/754,641

PRIOR FILING DATE: 2004-01-12

PRIOR APPLICATION NUMBER: 09/479,877

PRIOR FILING DATE: 2000-01-10

PRIOR APPLICATION NUMBER: 08/789,145

PRIOR FILING DATE: 1997-01-24

PRIOR APPLICATION NUMBER: 08/243,482

PRIOR FILING DATE: 1994-05-13

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patent version 3.2

SEQ ID NO: 9

LENGTH: 136

TYPE: PR

ORGANISM: Escherichia coli

US-10-754-641-9

Query Match 100.0%; Score 699; DB 5; Length 136;

Best Local Similarity 100.0%; Pred. No. 5e-67;

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 60  
1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 60

QY 61 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 120  
61 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 120

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

CITY: Fairfax  
STATE: VA  
COUNTRY: US  
ZIP: 22031  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/768,093  
FILING DATE: 02-Feb-2004  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,877B  
FILING DATE: 10-Jan-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenn M  
REGISTRATION NUMBER: 32,535  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703/425-8405  
TELEFAX: 703/425-8406  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 154 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: unknown  
TOPOLOGY: unknown  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-10-768-093-5

Query Match 100.0%; Score 699; DB 4; Length 154;  
Best Local Similarity 100.0%; Pred. No. 6e-67;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 60  
1 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 60

QY 19 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 78  
19 RTEIATKNPFVSTTISKSPFAPEPRIOPSFGENVGEGALLFSVNLTPBENTSOVTVVYV 78

QY 61 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 120  
61 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 120

QY 79 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 138  
79 YDEDYGLRLVNTADASQSIYIYVDEKGMKLDHGAETVTPNOQITFFALNVTSGEKKI 138

QY 121 SPGIYNDQVMVGYVYN 136  
121 SPGIYNDQVMVGYVYN 136

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

QY 139 SPGIYNDQVMVGYVYN 154  
139 SPGIYNDQVMVGYVYN 154

LENGTH: 154  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-754-641-5

Query Match 100.0%; Score 699; DB 5; Length 154;  
Best Local Similarity 100.0%; Pred. No. 6e-67;  
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RTEATKPFVSTTSSKSPFAPEPRIOPSGFENGKEGALLFSVNLTVPENVSQVTVYP 60  
DB 19 RTEATKPFVSTTSSKSPFAPEPRIOPSGFENGKEGALLFSVNLTVPENVSQVTVYP 78  
QY 61 YDEBDYGLRLVNTADASQSIYQIVDEKGGKMLKDGHAEVTPNQOITFKALNTSGEKXI 120  
DB 79 YDEBDYGLRLVNTADASQSIYQIVDEKGGKMLKDGHAEVTPNQOITFKALNTSGEKXI 138  
QY 121 SPGIYNDQVMVGYVYN 136  
DB 139 SPGIYNDQVMVGYVYN 154

## RESULT 5

US-10-768-093-10  
Sequence 10, Application US/10768093  
Publication No. US20040156829A1

## GENERAL INFORMATION:

APPLICANT: Wolf, Marcia K

Cassels, Fred J

Boedeker, Edgar C

TITLE OF INVENTION: Transformed Bacteria Producing G86

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hendricks and Assoc

STREET: P.O. Box 2509

CITY: Fairfax

STATE: VA

COUNTRY: US

ZIP: 22031

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/768,093

FILING DATE: 02-Feb-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/479,877B

FILING DATE: 10-Jan-2000

ATTORNEY/AGENT INFORMATION:

NAME: Hendricks, Glenna M

REGISTRATION NUMBER: 32,535

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703/425-8405

TELEFAX: 703/425-8406

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-768-093-10

Query Match 16.2%; Score 113.5; DB 4; Length 146;  
Best Local Similarity 28.5%; Pred. No. 0.00061;  
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

QY 7 KNFPVSTTSSKSPFAPE---PRIOPSGFENGKEGALLFSVNLTVPENVSQVTVYP--- 59  
DB 6 KSLDVNVIENQNF-LPIDSARIIIPVNYDSDPKLDSQLYTVEMTIPAGVSAVXIAPIDS 64

QY 60 YDEBDYGLRLVNTADASQSIYQIVDEKGGKMLKDGHAEVTPNQOITFKALNTSGEKXI 108  
DB 65 LTSSGQOIGKLVNANNPDQNNNYII-----RKDSAGNFMAGQKSPFKENTSYTF 116

QY 109 KALNTSGE---KKISPGIYNDQVMVGYVYN 136  
DB 117 SAI-YTGEIYNSGYSSTGTAGNLTGSEYSN 146

## RESULT 6

US-10-768-093-6  
Sequence 6, Application US/10768093  
Publication No. US20040156829A1

## GENERAL INFORMATION:

APPLICANT: Wolf, Marcia K

Cassels, Fred J

Boedeker, Edgar C

TITLE OF INVENTION: Transformed Bacteria Producing G86

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hendricks and Assoc

STREET: P.O. Box 2509

CITY: Fairfax

STATE: VA

COUNTRY: US

ZIP: 22031

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/768,093

FILING DATE: 02-Feb-2004

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/479,877B

FILING DATE: 10-Jan-2000

ATTORNEY/AGENT INFORMATION:

NAME: Hendricks, Glenna M

REGISTRATION NUMBER: 32,535

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703/425-8405

TELEFAX: 703/425-8406

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 167 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-10-768-093-6

Query Match 16.2%; Score 113.5; DB 4; Length 167;  
Best Local Similarity 28.5%; Pred. No. 0.00074;  
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

QY 7 KNFPVSTTSSKSPFAPE---PRIOPSGFENGKEGALLFSVNLTVPENVSQVTVYP--- 59  
DB 27 KSLDVNVIENQNF-LPIDSARIIIPVNYDSDPKLDSQLYTVEMTIPAGVSAVXIAPIDS 85

QY 60 YDEBDYGLRLVNTADASQSIYQIVDEKGGKMLKDGHAEVTPNQOITFKALNTSGEKXI 108  
DB 86 LTSSGQOIGKLVNANNPDQNNNYII-----RKDSAGNFMAGQKSPFKENTSYTF 137

QY 109 KALNTYSGE---KXISPGIYNDQVWVGYYN 136  
DB 138 SAI-YTGGEPYNSGYSSTGYAGNLTVSFYSN 167

RESULT 7  
US-10-754-641-6  
Sequence 6, Application US/10754641  
Publication No. US20050025787A1

GENERAL INFORMATION:  
APPLICANT: U.S. Army Medical Research and Materiel Command  
APPLICANT: Wolf, Marcia K.  
APPLICANT: Cassele, Frederick J.  
APPLICANT: Boedeker, Edgar C.  
TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES  
FILE REFERENCE: 034047.033.3  
CURRENT APPLICATION NUMBER: US/10/754,641  
CURRENT FILING DATE: 2004-01-12  
PRIOR APPLICATION NUMBER: 09/479,877  
PRIOR FILING DATE: 2000-01-10  
PRIOR APPLICATION NUMBER: 08/788,145  
PRIOR FILING DATE: 1997-01-24  
PRIOR APPLICATION NUMBER: 08/243,482  
PRIOR FILING DATE: 1994-05-13  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 6  
LENGTH: 167  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-754-641-6

Query Match 16.2%; Score 113.5; DB 5; Length 167;  
Best Local Similarity 28.5%; Pred. No. 0.00074;  
Matches 43; Conservative 22; Mismatches 55; Indels 31; Gaps 7;

QY 7 KNFVSTTISKSPFAPE---PRIQSFGENYKEGALLFSVNLTVPENVSQVYTP--- 59  
DB 27 KSLDVNVNIEQNFI-PDIDSAVRIFPNYDSPKLDLSQLYTEMTIPAGVSAVKIAPLDS 85  
QY 60 VYDEYDGLRYNTADASQSIYQYVDEKGMKLDHGA-----EYTPNQDITF 108  
DB 86 LTSSGOQIGKLVNPNPDQNMNYI-----RKDSGAGNFMAGOKGSFVYKENTSYTF 137  
QY 109 KALNTYSGE---KXISPGIYNDQVWVGYYN 136  
DB 138 SAI-YTGGEPYNSGYSSTGYAGNLTVSFYSN 167

RESULT 8  
US-10-267-989-25  
Sequence 25, Application US/10367989  
Publication No. US20030199035A1  
GENERAL INFORMATION:  
APPLICANT: ARNAU, Jose  
APPLICANT: VRANG, Astrid  
APPLICANT: ISRAELSEN, Hans  
APPLICANT: JOERGENSEN, Flemming  
APPLICANT: MADSEN, Soeren  
TITLE OF INVENTION: METABOLICALLY ENGINEERED LACTIC ACID BACTERIA AND  
TITLE OF INVENTION: MEANS FOR PROVIDING  
FILE REFERENCE: ARNAU-1A  
CURRENT APPLICATION NUMBER: US/10/267,989  
CURRENT FILING DATE: 2002-10-10  
PRIOR APPLICATION NUMBER: US/08/981,097  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: PCT/DK97/00336  
PRIOR FILING DATE: 1997-08-20  
PRIOR APPLICATION NUMBER: 08/701,458  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn version 3.0  
SEQ ID NO 25  
LENGTH: 740  
TYPE: PRT  
ORGANISM: Clostridium pasteurianum  
US-10-267-989-25

Query Match 11.1%; Score 77.5; DB 4; Length 740;  
Best Local Similarity 25.0%; Pred. No. 44;  
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

QY 4 IATKN-PVSTTISKSPFAPEPRIQSFGENYKEGALLFSVNLTVPENVSQV---TVY 58  
DB 336 LVTKNSFRYHLTLINLSGAPBNMTVMSENV-PBSFKKFAEHSITLDSIQYENDDIR 394  
QY 59 PUYDEDYGL-----GRLVNTADASQSIYQYVDEKGMKLDHGAETP 102  
DB 395 PIYGDVAILACCVSAMRVGKMQFFGARCNIKLAKCLLAINGVDEK-----KXIKVVP 447  
QY 103 N-QQITFPALNTSGEKKISPGIYND-QVWVGYYN 136  
DB 448 DIEPTDEVLDY---EKVKENYFVLEVMAGLYVN 479

RESULT 9  
US-10-357-567-25  
Sequence 25, Application US/10357567  
Publication No. US20040038382A1  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter  
TITLE OF INVENTION: Compositions and Methods for Bioremediation  
FILE REFERENCE: OHU-07748  
CURRENT APPLICATION NUMBER: US/10/357,567  
CURRENT FILING DATE: 2003-02-04  
PRIOR APPLICATION NUMBER: 09/072,433  
PRIOR FILING DATE: 1998-05-04  
PRIOR APPLICATION NUMBER: 60/046,845  
PRIOR FILING DATE: 1997-05-05  
NUMBER OF SEQ ID NOS: 71  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 25  
LENGTH: 740  
TYPE: PRT  
ORGANISM: Clostridium pasteurianum  
US-10-357-567-25

Query Match 11.1%; Score 77.5; DB 4; Length 740;  
Best Local Similarity 25.0%; Pred. No. 44;  
Matches 39; Conservative 22; Mismatches 60; Indels 35; Gaps 8;

QY 4 IATKN-PVSTTISKSPFAPEPRIQSFGENYKEGALLFSVNLTVPENVSQV---TVY 58  
DB 336 LVTKNSFRYHLTLINLSGAPBNMTVMSENV-PBSFKKFAEHSITLDSIQYENDDIR 394  
QY 59 PUYDEDYGL-----GRLVNTADASQSIYQYVDEKGMKLDHGAETP 102  
DB 395 PIYGDVAILACCVSAMRVGKMQFFGARCNIKLAKCLLAINGVDEK-----KXIKVVP 447  
QY 103 N-QQITFPALNTSGEKKISPGIYND-QVWVGYYN 136  
DB 448 DIEPTDEVLDY---EKVKENYFVLEVMAGLYVN 479

RESULT 10  
US-10-369-493-21384  
Sequence 21384, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng

TYPE: prp  
ORGANISM: *Pasteurella multocida*

```

US-10-282-122A-67315
Query Match      10.8%; Score 75.5; DB 4; Length 246;
Best Local Similarity 21.4%; Pred. No. 16;
Matches 28; Conservative 25; Mismatches 49; Indels 29; Gaps 5;

QY 9 FVSTTISKSFFAPBPRI--QPSFGENVKKGALLFVNLTVPENVSVQTVTPYD---- 62
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 111 FGIIGLIDVASYSPPLKRVENOSRFGDTLGRYV-----ETGSYMLPLYGPAAP 159
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 63 -EDYGLGRLVNTADASQSI-----YQIVDEKGMKLDKHGAETVPNOQITFKALN 112
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 160 ROD--IGNLADTTYPMLSLGGWMLSKSGVQIIDGRKALDQALLGOSQDPYITFREAY 217
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 113 YTSGEKKSIPG 123
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 218 FQNLERYVMDG 228

RESULT 13
US-10-282-122A-44691
/ Sequence 44691, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zykend, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Foreyeth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See file wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 44691
/ LENGTH: 640
/ TYPE: PRT
/ ORGANISM: Acinetobacter baumannii
US-10-282-122A-44691
Query Match      10.7%; Score 75; DB 4; Length 640;
Best Local Similarity 26.0%; Pred. No. 67;
Matches 26; Conservative 19; Mismatches 47; Indels 8; Gaps 2;

```

```

QY 26 IQPSFGENVKKGALLFVNLTVPENVSVQTV-----VYFVDEDIYGLRVNTADASQ 78
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 64 VQGLDEVARVAKLTPTDFSLKKGSEEVKTLTSLSEARVERCMAEKY-RGSINMKVASQ 122
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 79 SIYQIVDEKGMKLDKHGAETVPNOQITFKALNYTSGEK 118
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 123 FFIDELIDDKYQKLVKAKASDEITQHLMLPMLALSSQE 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14
US-09-815-242-11293
/ Sequence 11293, Application US/09815242
/ Patent No. US20020061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari L.
/ APPLICANT: Zykend, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11293
/ LENGTH: 736
/ TYPE: PRT
/ ORGANISM: Helicobacter pylori
US-09-815-242-11293
Query Match      10.7%; Score 75; DB 3; Length 736;
Best Local Similarity 19.7%; Pred. No. 82;
Matches 23; Conservative 29; Mismatches 63; Indels 2; Gaps 1;

QY 1 RTEIATKRPVSTTISKSFFAPBPRIOPSPFGENVKKGALLFVNLTVPENVSVQTVVPV 60
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 228 KNELEKESYIAISSVYKSKSPTP--PPFMTSTIQQSASSLIGSPTTMTIAQULYSGV 285
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 YDEYGLGRLVNTADASQSIYQIVDEKGMKLDKHGAETVPNOQITFKALNYTSGE 117
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 286 ATPGVWGVIYTRMTDSINIKALEARNKILDKYGDYVPPRAKYVSSXNKVQAOE 342
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
US-10-282-122A-58624
/ Sequence 58624, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zykend, Judith

```



APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Foreyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See file wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 58624  
LENGTH: 736  
TYPE: PRT  
ORGANISM: Helicobacter pylori  
US-10-282-122A-58624

Query Match 10.7%; Score 75; DB 4; Length 736;  
Best Local Similarity 19.7%; Pred. No. 82;  
Matches 23; Conservative 29; Mismatches 63; Indels 2; Gaps 1;

OY 1 RTETATKNPVSITSTKSFPAPEPRIOPSFGENVGEGALLFSVNLTVPENVSQVTVYPV 60  
DB 228 KNELEKESEYALSSIVKSKKSPTP--PPFMTSTLQGSASSLGFSPPTKMSIAQKLYEGV 285  
OY 61 YDEHYGLGRLVNTADASQSIYQIVDEKKGKMLKDHALEVTNPNOQITFKALNVTSGE 117  
DB 286 ATPGVMGVITYMRTDSINIAKEALEARNKILKDYLGKDYLPKAKVYSSIKNKNAOE 342

Search completed: April 10, 2006, 10:25:12  
Job time : 167 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 10, 2006, 10:22:35 ; Search time 25 Seconds

(without alignments)  
169.684 Million cell updates/sec

Title: US-10-768-093-9

Perfect score: 699  
Sequence: 1 RTEATKXNPVSTTRISKSF.....EKKISPGIYDQWVGYVYN 136

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 184161 seqs, 31191982 residues

Total number of hits satisfying chosen parameters: 184161

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_New.\*  
1: /SIDS5/prodata/2/pubppa/US08\_NEW\_PUB pep.\*  
2: /SIDS5/prodata/2/pubppa/US07\_NEW\_PUB pep.\*  
3: /SIDS5/prodata/2/pubppa/US07\_NEW\_PUB pep.\*  
4: /SIDS5/prodata/2/pubppa/PCF\_NEW\_PUB pep.\*  
5: /SIDS5/prodata/2/pubppa/US05\_NEW\_PUB pep.\*  
6: /SIDS5/prodata/2/pubppa/US10\_NEW\_PUB pep.\*  
7: /SIDS5/prodata/2/pubppa/US11\_NEW\_PUB pep.\*  
8: /SIDS5/prodata/2/pubppa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	11.2	268	US-11-195-739-6	Sequence 6, Appl 1
2	75.5	10.8	356	US-11-096-568A-32647	Sequence 32647, A
3	75.5	10.8	361	US-11-096-568A-32646	Sequence 32646, A
4	74.5	10.7	400	US-10-793-626-2774	Sequence 2774, Ap
5	72.5	10.4	375	US-10-485-913-166	Sequence 166, App
6	71.5	10.2	1680	US-10-517-939-362	Sequence 362, App
7	70.5	10.1	357	US-11-194-246-282	Sequence 282, App
8	70.5	10.1	898	US-11-099-691-7	Sequence 7, Appl 1
9	70.5	10.1	898	US-11-124-367A-330	Sequence 330, App
10	70.5	10.1	898	US-11-124-367A-333	Sequence 333, App
11	69.5	9.9	379	US-11-096-568A-6312	Sequence 6312, Ap
12	69.5	9.9	416	US-11-096-568A-6311	Sequence 6311, Ap
13	69.5	9.9	450	US-11-096-568A-6310	Sequence 6310, Ap
14	69	9.9	501	US-11-087-099-5465	Sequence 5465, Ap
15	68.5	9.8	278	US-10-793-626-2208	Sequence 2208, Ap
16	68.5	9.8	328	US-10-506-454-335	Sequence 335, App
17	68.5	9.8	340	US-11-057-720-1372	Sequence 52, Appl
18	68.5	9.8	357	US-09-895-493-178	Sequence 178, App
19	68.5	9.8	616	US-11-083-800-4	Sequence 4, Appl 1
20	68	9.7	299	US-11-156-084-339	Sequence 339, App
21	68	9.7	357	US-11-087-099-5465	Sequence 5465, Ap
22	68	9.7	357	US-11-087-099-7487	Sequence 7487, Ap
23	68	9.7	384	US-10-517-939-358	Sequence 358, App
24	68	9.7	386	US-11-096-568A-18911	Sequence 18911, A
25	68	9.7	415	US-11-096-568A-18910	Sequence 18910, A

26	68	9.7	452	US-11-096-568A-18909	Sequence 18909, A
27	68	9.7	501	US-11-087-099-5784	Sequence 5784, Ap
28	68	9.7	501	US-11-087-099-8619	Sequence 8619, Ap
29	67.5	9.7	1377	US-10-467-657-7922	Sequence 7922, Ap
30	67	9.6	273	US-11-051-720-1375	Sequence 1375, Ap
31	67	9.6	729	US-11-051-720-1373	Sequence 1373, Ap
32	67	9.6	738	US-11-051-720-1374	Sequence 1374, Ap
33	67	9.6	1081	US-11-051-720-1372	Sequence 1372, Ap
34	67	9.6	1767	US-10-995-561-911	Sequence 911, App
35	67	9.6	1767	US-10-995-561-914	Sequence 914, App
36	67	9.6	1806	US-10-995-561-912	Sequence 912, App
37	67	9.6	1806	US-10-995-561-915	Sequence 915, App
38	67	9.6	1806	US-11-051-720-1446	Sequence 1446, Ap
39	67	9.6	1806	US-11-051-720-1447	Sequence 1447, Ap
40	67	9.6	1818	US-10-995-561-910	Sequence 910, App
41	67	9.6	1818	US-10-995-561-913	Sequence 913, App
42	66.5	9.5	355	US-10-517-939-164	Sequence 164, App
43	66	9.4	448	US-10-506-454-501	Sequence 501, App
44	66	9.4	961	US-11-051-724-42	Sequence 42, Appl
45	66	9.4	1874	US-10-821-234-1182	Sequence 1182, Ap

## ALIGNMENTS

```
RESULT 1
US-11-195-739-6
; Sequence 6, Application US/11195739
; Publication No. US20060019324A1
;
GENERAL INFORMATION:
; APPLICANT: ALZARI, PEDRO
; APPLICANT: BOITTEL, BIRGITTE
; APPLICANT: VILLARINO, ANDREA
; APPLICANT: FERNANDEZ, PABLO
; APPLICANT: COLE, STEWART
;
TITLE OF INVENTION: PKB KINASE AND P38 MAPKINASE AND METHODS OF IDENTIFYING
; TITLE OF INVENTION: INHIBITORY SUBSTANCES
;
FILE REFERENCE: 25285305
;
CURRENT APPLICATION NUMBER: US/11/195,739
;
CURRENT FILING DATE: 2005-08-03
;
PRIOR APPLICATION NUMBER: US/10/892,170
;
PRIOR FILING DATE: 2004-07-16
;
PRIOR APPLICATION NUMBER: US 60/487,943
;
NUMBER OF SEQ ID NOS: 20
;
SOFTWARE: PatentIn version 3.2
;
SEQ ID NO 6
;
LENGTH: 268
;
TYPE: PRT
;
ORGANISM: Thermobifida fusca
;
US-11-195-739-6
;
Query Match
Best Local Similarity 26.4%; Pred. No. 1.4; Length 268;
Matches 39; Conservative 22; Mismatches 43; Indels 44; Gaps 8;
;
2 TEATKXNP-----VSTTRISKFPAPRPIOPSGENVKAGALLFVNLTVPENV 52
112 TWITTKPAPGEANREBSVTLTISAGF--PMPNV---VGQKVDARRLLESDDL----- 160
;
;
53 SOWTYPPYDEDDYGLGLVN-----TADASQST-----ITYQVLD-----EKCK 90
;
;
161 EYTVBEHHEHVEPGEHVISOPEAETTVGAGOSVTLTVSSGPELVEVFDIRGMKVDKAR 219
;
;
91 KMLKHGABVTNNOQITFKALNYT-SGE 117
220 KELEBGEVTVHGVIGKRVGDYMPKGE 247
;
;
RESULT 2
US-11-096-568A-32647
; Sequence 32647, Application US/11096568A
; Publication No. US20060048240A1
```

RESULT 4  
US-10-793-626-274 :  
; Sequence 2774, Application US/10793626  
; Publication No. US2005025478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMBERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

RESULT 6  
US-10-517-939-362

```

; Sequence 362, Application US/10517939
; Publication No. US20060003433a1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esterhalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 56446207901
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: US/10/517,939
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: ParseSeq for Windows Version 4.0
; SEQ ID NO 362
; LENGTH: 1680
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample.
; NAME/KEY: SIGNAL
; LOCATION: (1)...(26)
; US-10-517-939-362

Query Match      10.1%; Score 71.5; DB 6; Length 1680;
Best Local Similarity 21.3%; Pred. No. 84;
Matches 37; Conservative 27; Mismatches 55; Indels 55; Gaps 8;

Qy      1 RTEI--ATKNPVSTTISKSFAPRIQPSFGENVGKGLLPSVNI-TVPEVSOVTV 57
      1032 RTNVAPATVTTIDVTL-----RQINTLNAAGRGRIQTGNAPTITIDVFEV 1083
      58 YPVYVDE-----YGLGR-----LVNTADASQSIIVQ---I 84
      1084 YQIDIDIDPAGLPPEQWNPDLPRISELFEPYFGLNIVSTETLMANETKQAFHHEVNI 1143
      85 VDEKSKMLKDHGAEVT-----PNOQITFKALNTSGEKISPGIYNDQWVGVY 133
      1144 TAENGHKSSIAGPENSTFVPEPEQ-----FNFTDADRIYVFAVENDIELVGH 1191
      Db

RESULT 7
US-11-194-246-282
; Sequence 282, Application US/11194246
; Publication No. US20050272089a1
; GENERAL INFORMATION:
; APPLICANT: Mott, John
; APPLICANT: Trepod, Catherine
; APPLICANT: Arvidson, Stefan
; TITLE OF INVENTION: CRITICAL GENES AND POLYPEPTIDES OF HAEMOPHILUS INFLUENZAE AND MET
; TITLE OF INVENTION: USE
; FILE REFERENCE: 00592.US1 (MAR 268.05920101)
; CURRENT FILING DATE: 2005-08-01
; PRIOR APPLICATION NUMBER: US/11/194,246
; PRIOR FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US 60/345,438
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 621
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 282
; LENGTH: 357
; TYPE: PRT
; ORGANISM: HAEMOPHILUS INFLUENZAE
; US-11-194-246-282

```

```

Query Match      10.1%; Score 70.5; DB 7; Length 357;
Best Local Similarity 25.8%; Pred. No. 13;
Matches 32; Conservative 23; Mismatches 50; Indels 19; Gaps 7;

Qy      16 SKGFAPPEPRIQSPGR---NVGKGLLPSVNLTVPEVNSQTVVYVYDE-DYGLRLV 71
      180 SNPPCPDESVAVERFDLIRELKKGSDISGAKLTVIENVPVGLGEVFPRLDADLHAL 239
      72 NTADASQSII-----YQIVDEKSKMLKDHGAEVTNPQOITFKALNTSGEKISPGIYND 127
      240 MGINAVGVAVIGGFAVNEGRG-----SEHDEMTPN---GFES-NHAGC---ILGGISSG 288
      Qy      128 QWNV 131
      Db      289 QPIL 292

RESULT 8
US-11-099-691-7
; Sequence 7, Application US/11099691
; Publication No. US20050260644a1
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: BANDMAN, Olga
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YANG, Junning
; TITLE OF INVENTION: CELL SIGNALING PROTEINS
; FILE REFERENCE: PE-0521 PCT
; CURRENT APPLICATION NUMBER: US/11/099,691
; CURRENT FILING DATE: 2005-04-06
; PRIOR APPLICATION NUMBER: US/09/700,444
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/085,343
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,010
; PRIOR FILING DATE: 1998-08-26
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 898
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte Clone 2948818
; US-11-099-691-7

Query Match      10.1%; Score 70.5; DB 7; Length 898;
Best Local Similarity 23.2%; Pred. No. 45;
Matches 23; Conservative 26; Mismatches 27; Indels 23; Gaps 5;

Qy      28 PSFGENVGKGLLPSVNLTVPEVNSQTVVYVYDEYGLRLVNTADASQSIIVQ---I 83
      351 PSMVSVTSSQSSSVNSQWVWDESSSELVW--WMDDE-----STINSSSVVHKKHCHV 401
      84 -TVDEKSKMLKDHG---AEVTPNQOITFKALNTSGEK 118
      402 FIRDEAG-----HGDRPEPRPTQSVOSQALHYRNRR 434
      Db

RESULT 9
US-11-124-367A-330
; Sequence 330, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; US-11-124-367A-330

```

RESULT 13  
US-11-096-568A-6310  
; Sequence 6310, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 10, 2006, 09:55:20 ; Search time 26.7 Seconds  
(without alignments)  
1734.477 Million cell updates/sec

Title: US-10-768-093-4

Perfect score: 939  
Sequence: 1 TTGACACATTACGAATGTTA.....TTGGTTACTACTGTAATACTAA 527

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p model -DEV=xlp  
-DB=ABSSWAB\_pool/US10768093/runac\_10042006\_090301\_4503/APP\_QUERY.fasta.1  
-DB=GeneSeq -QFMT=fastac -SUFFIX=n2p.rag -WIND=100 -I -LOOPEL=0 -LOOPEXT=0  
-UNITS=dits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs802p  
-USER=us10768093 @GCN 1.1 348 @runac 10042006 090301 4503 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_GeneSeq\_21.1  
1: geneeqp19808:\*  
2: geneeqp19908:\*  
3: geneeqp20008:\*  
4: geneeqp20018:\*  
5: geneeqp20028:\*  
6: geneeqp20038:\*  
7: geneeqp20048:\*  
8: geneeqp20058:\*  
9: geneeqp20068:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785	83.6	154	8	ADRI4912 Escherich
2	785	83.6	154	8	ADRI4912 Escherich
3	699	74.4	136	8	ADRI4916
4	699	74.4	136	8	ADRI4916
5	118	12.6	167	8	ADRI4913 Escherich
6	118	12.6	167	8	ADRI4913 Escherich
7	113.5	12.1	146	8	ADRI4917 CS6 relat
8	92.5	9.9	221	6	AAE37981 Human kin
9	90.5	9.6	268	9	ADWA7927 Thermobif

10	87.5	9.3	911	6	ABR41065	ADRI41065 Human MAP
11	87.5	9.3	911	7	ADV68432	ADV68432 Human HIP
12	86	9.2	378	7	ADC95115	ADC95115 E. faeciu
13	85	9.1	506	4	ABR49208	ABR49208 V.cholera
14	85	9.1	541	6	ABU49234	ABU49234 Protein e
15	84.5	9.0	1239	9	ADR86415	ADR86415 Aspergill
16	84	8.9	356	9	ABR40460	ABR40460 L. pneumo
17	84	8.9	384	9	ABR37118	ABR37118 L. pneumo
18	84	8.9	439	8	ADN19836	ADN19836 Bacteria
19	83	8.8	183	5	ABR49515	ABR49515 listeria
20	83	8.8	750	6	ADB09635	ADB09635 Altiococ
21	82	8.7	610	9	ABR37679	ABR37679 L. pneumo
22	82	8.7	610	9	ABR40990	ABR40990 L. pneumo
23	81	8.6	768	7	ADH88582	ADH88582 Enterococ
24	79.5	8.5	422	5	ADP28400	ADP28400 Bacteria
25	79.5	8.5	835	5	ABP26319	ABP26319 Streptoc
26	79.5	8.5	951	6	ABR41062	ABR41062 Human MAP
27	79.5	8.5	1171	4	ABR65661	ABR65661 Novel pro
28	79.5	8.5	1171	5	ABP69278	ABP69278 Human pol
29	79.5	8.5	1171	5	AAE16262	AAE16262 Human kin
30	79.5	8.5	1171	6	ABR41063	ABR41063 Human MAP
31	79.5	8.5	1171	8	ADP28268	ADP28268 Human PRO
32	79.5	8.5	1171	9	ADY18325	ADY18325 PRO polyp
33	79.5	8.5	1171	9	ADP69277	ADP69277 Human pol
34	79.5	8.5	1198	5	ABP69277	ABP69277 Human pol
35	78.5	8.4	332	3	AAE10781	AAE10781 Arabidops
36	78.5	8.4	355	3	AAE10780	AAE10780 Arabidops
37	78.5	8.4	355	3	ABR36463	ABR36463 A. thalia
38	78.5	8.4	367	3	AAE10779	AAE10779 Arabidops
39	78.5	8.4	989	9	ADV68430	ADV68430 Human hom
40	78.5	8.4	1191	5	AAO18296	AAO18296 Human hom
41	78.5	8.4	1198	5	ABG61938	ABG61938 Protease
42	78.5	8.4	1198	5	AAO22555	AAO22555 Serine-ri
43	78.5	8.4	1198	5	ABG73319	ABG73319 Human kin
44	78.5	8.4	1198	6	ABR41064	ABR41064 Human MAP
45	78.5	8.4	1198	7	ADD18976	ADD18976 Human dis

## ALIGNMENTS

RESULT 1	ADRI4912	standard; protein; 154 AA.
XX	ADRI4912;	
AC	ADRI4912;	
DT	XX	
XX	XX	
DE	04-NOV-2004 (first entry)	
XX	Escherichia coli E8775 antigen csaA protein.	
XX	CS6; csaA; csaB; csaC; csaD; origin of replication; lac promoter;	
KW	kanamycin resistance gene; CS6a; CS6b; enterotoxigenic E. coli;	
KW	Escherichia coli E8775.	
XX	XX	
OS	Escherichia coli.	
XX	XX	
PN	US2004156829-A1.	
XX	XX	
PD	12-AUG-2004.	
XX	XX	
PF	02-FEB-2004; 2004US-00768093.	
XX	XX	
PR	13-MAY-1994; 94US-0023482.	
PR	24-JAN-1997; 97US-00788145.	
XX	10-JAN-2000; 2000US-00479877.	
XX	XX	
PA	(USSA) US SEC OF ARMY.	
XX	XX	
PI	Wolf MK. Cassels FU, Boedeker EC.	
XX	XX	
DR	WPI; 2004-592722/57.	
XX	N-PSDB; ADRI4911.	
XX	XX	







XX PS Claim 4; SEQ ID NO 9; 23pp; English.

CC The invention relates to a method of inducing in a susceptible host, the

CC production of antibodies against CS6 protein comprising giving a

CC composition of matter made of bacteria transformed with a plasmid, which

CC contain genes *csaA* and *csaB*, all of *csaC* and DNA sequence *csaD* that

CC encodes at least 802 amino acids (at least 2406 base pairs), an origin of

CC replication, a *lac* promoter, and a kanamycin resistance gene, where the

CC bacteria overexpress both *CS6A* and *CS6B* proteins. The method and proteins

CC are useful for stimulating protective antibodies against enterotoxigenic

CC *Escherichia coli*. This sequence corresponds to the *E. coli* *CS6 csaA*

CC mature protein

XX SQ Sequence 136 AA;

Alignment Scores:

Pred. No.:	2,92e-82	Length:	136
Score:	699.00	Matches:	136
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	74.4%	Indels:	0
DB:	9	Gaps:	0

US-10-768-093-4 (1-527) x ADW47737 (1-136)

OY 117 AGAACAAGAAATAGCGACTAAACCTCCAGATCAACGACTATTTCAAAAAGTTT 176

DB 1 ATGTHGIIILALELHRYSAENPHEPROVALSERTHRILESERYSERPHEPE 20

OY 177 GCACCTGAACCAAGCAATACAGCTTTTGGTGGAAAAAGTTGGAAGAGAGCTTTA 236

DB 21 AIAPIROGLIUPROARGLIEGINPROSERPHEGLIUAENVAGLYLSEGLIUALALEU 40

OY 237 TTAATTAGTGAAGTAACTGTTCTGAAAGTATCCAGGTACGGCTACCTGTT 296

DB 41 LEUPHSEVERVALSNEUTHRVAILPROGLUENVALSERGLINVAITHRVALTYRPROVAL 60

OY 297 TATGATGAAGATTATGCTTAGACGACACTAGTAATATACCGCTGATCTTCCCATCAATA 356

DB 61 TYRAPSGIUAEPYICLYLEUGIYARGLEUVALAENTHRLAASPILASERGINSEITILE 80

OY 357 ATCTACCAATGTTGATGAGAAAGGAAAAAATGTTAAAGATCATGTCGACGCTT 416

DB 81 ILETYRGIIILEVALAEPGIULYSGLIYLYSMETLEULYSEPHISEGLIYALAGIULVAL 100

OY 417 AACACCTAATCAACAATACTTTAAAGCGCTAATTATCTAGCGGGAAAAAATAA 476

DB 101 THIRPROAENGILNILETHIRPHELYSALAEUASNTYRTHSERGLYGLIULYSYLE 120

OY 477 TCTCCTGAATATATAACGATCAGTTAGCTTGTGTTACTATGTAAC 524

DB 121 SERPROGLIYILETYRSHNAPGILNVALMETVALGITYRTYRVALASN 136

RESULT 5

ID ADR14913 standard; protein; 167 AA.

XX ADR14913,

XX 04-NOV-2004 (first entry)

DE *Escherichia coli* E8775 antigen *csaB* protein.

KM *CS6*, *csaA*, *csaB*, *csaC*, *csaD*; origin of replication; *lac* promoter;

KM kanamycin resistance gene; *CS6A*, *CS6B*; enterotoxigenic *E. coli*;

KM *Escherichia coli* E8775.

XX *Escherichia coli*.

XX US2004156829-A1.

XX 12-AUG-2004.

XX PF 02-FEB-2004; 2004US-00768093.

XX PR 13-MAY-1994; 94US-00243482.

XX PR 24-JAN-1997; 97US-00788145.

XX PR 10-JAN-2000; 2000US-00479877.

XX (USSA ) US SEC OF ARMY.

XX Wolf MK, Casele's FJ, Boedeker EC,

XX WPI; 2004-592722/57.

XX DR

XX PT Inducing the production of antibodies against CS6 protein, useful for

XX PT preventing pathological effects of enterotoxigenic *E. coli*, by

XX PT administering a composition comprising transformed bacteria producing CS6

XX PT antigens.

XX PS Disclosure; SEQ ID NO 6; 22pp; English.

CC The invention relates to a method of inducing, in a susceptible host, the

CC production of antibodies against a CS6 protein, comprising administering

CC a composition of matter comprising bacteria transformed with a plasmid

CC containing genes *csaA*, *csaB*, *csaC* and *csaD*, an origin of replication, a

CC *lac* promoter and a kanamycin resistance gene, where the bacteria

CC expresses both *CS6A* and *CS6B* proteins. The invention also relates to a

CC composition of matter comprising a protein in a pharmaceutical carrier.

CC The pharmaceutical carrier is a carbonated liquid. The method is useful

CC for inducing the production of antibodies against the CS6 protein. The

CC protein and composition are used as vaccines for preventing pathological

CC effects of enterotoxigenic *E. coli*. This sequence represents the

CC *Escherichia coli* E8775 antigen *csaB* of the invention.

XX SQ Sequence 167 AA;

Alignment Scores:

Pred. No.:	7,74e-06	Length:	167
Score:	118.00	Matches:	49
Percent Similarity:	42.1%	Conservative:	26
Best Local Similarity:	27.5%	Mismatches:	67
Query Match:	12.6%	Indels:	36
DB:	8	Gaps:	9

US-10-768-093-4 (1-527) x ADR14913 (1-167)

OY 63 ATGAAGAAACAAATGTTTAAATCTTAATCTTGTCTTATCGGACCATCCAGAA 122

DB 2 LEUYLSYLSILELLESERLALILELLEULEIA-----GLYTHSERGLYVALVAL 19

OY 123 GAATATGCGACT-----AAAACCTCCAGATCAACGACTATTTCAAAAAGTTT 173

DB 20 ASNALAGLYASNTYRGILNTYRYSERLEUASPVALAENVALSNILGLINDINAPNE 39

OY 174 TTGGAAGCTGA-----CCAGAAATACAGCTTCTTTGGTGAANAATGTGA 221

DB 40 ILE---PROASPILEASPSERLAVALARGILELLEPROVALASNTYRASPSESPRO 58

OY 222 AAGGAAGAGCTTTATTTATTTAGTGAATTAAGCTTCTTCAAAAATGATCCAGGTA 281

DB 59 LYSLEUASPSESGILNEUTYRTHVALGLUETHIRILEPROVALGLYVALSERLALVAL 78

OY 282 ACGGTCTACCT-----GTTATGATGAAGATTATGAGTTAGACGACTAGTAAT 332

DB 79 LYSILEALAPROTHRASPSERLHRYSESERGLYGLINILEGLYLSLEUVALASN 98

OY 333 ACCGTGATNGCTCCCAATCAATATATCAACGATGTTGATGAGAAAGGAAAAAATG 392

DB 99 VALAENAPROASPGILNASMETCASNTRYTYRILE-----GAGGTACA 419

OY 393 TTAAGAATCATGTGCA-----GAGGTACA 419

DB 111 ARGLYSAPSESGIYALAGLYASNPHMETCALAGLYGLINLYSGLYSERPHEPROVALLYS 130

OY 420 CCTAATCAACAAATTAAGCGCTGAATTAAGCGGGGAA-----AAA 470  
 DB 131 GluAenthrSerTyThrPheSerAlaIle---TyrThrGlyGlyGlyProAsnSer 149  
 OY 471 AAAATATCTCCTGGAATATATACATCAGCTTATGCTGTACTATGTAAC 524  
 DB 150 GlyTyrSerSerGlyThrTyraIaGlyAsnLeuThrValSerPheTySerAsn 167  
 RESULT 6  
 ADM47734  
 ID ADM47734 standard; protein; 167 AA.  
 XX ADM47734;  
 AC  
 XX 21-APR-2005 (first entry)  
 XX  
 DE E. coli CS6 ccsb full length protein.  
 XX  
 KM antibacterial; vaccine; DNA purification; colonization factor antigen.  
 XX  
 OS Escherichia coli.  
 XX  
 PN US2005025787-A1.  
 PD 03-FEB-2005.  
 XX  
 PF 12-JAN-2004; 2004US-00754641.  
 XX  
 PR 13-MAY-1994; 94US-00243482.  
 PR 24-JAN-1997; 97US-00788145.  
 PR 10-JAN-2000; 2000US-00479877.  
 XX  
 PA (WOLF/) WOLF M K.  
 PA (CASS/) CASSELS F J.  
 PA (BOED/) BOEDEKER E C.  
 PI Wolf MK, Casseels FJ, Boedeker EC;  
 PI WPI; 2005-131784/14.  
 DR  
 XX  
 PT Inducing in a susceptible host, the production of antibodies against CS6  
 PT protein, useful for stimulating protective antibodies against Escherichia  
 PT coli by administering a composition comprising bacteria transformed with  
 PT a plasmid.  
 PS Disclosure; SEQ ID NO 6; 23pp; English.  
 XX  
 CC The invention relates to a method of inducing in a susceptible host, the  
 CC production of antibodies against CS6 protein comprising giving a  
 CC composition of matter made of bacteria transformed with a plasmid, which  
 CC contain genes csaA and csaB, all of csaC and DNA sequence csaD that  
 CC encodes at least 802 amino acids (at least 2406 base pairs), an origin of  
 CC replication, a lac promoter, and a kanamycin resistance gene, where the  
 CC bacteria overexpress both CS6A and CS6B proteins. The method and proteins  
 CC are useful for stimulating protective antibodies against enterotoxigenic  
 CC Escherichia coli. This sequence corresponds to the E. coli CS6 ccsb  
 CC protein  
 XX  
 SO Sequence 167 AA;  
 Alignment Scores:  
 Pred. No.: 7.74e-06 Length: 167  
 Score: 118.00 Matches: 49  
 Percent Similarity: 42.16 Conservative: 26  
 Best Local Similarity: 27.54 Mismatches: 67  
 Query Match: 12.64 Indels: 36  
 DB: 9 Gaps: 9  
 US-10-768-093-4 (1-527). x ADM47734 (1-167).  
 OY 63 ATGAAGAAACAAATGGTTTATTCATCTTGGCTTATCGCAGCAGCATCCAGAAC 122  
 DB 2 LeuYelYelIleIleSerAlaIleAlaLeuIleA-----GlyThrSerGlyValVal 19

OY 123 GAAATAGGACT-----AAAACCTCCAGTATCAACGACTATTTCAAAAATT 173  
 DB 20 AsnAlaGlyAsnTpgInTyRlySerLeuAspValAsnIleGluGlnAsnDhe 39  
 OY 174 TTTCGACCTGAA-----CCAGGAATACAGCTTCTTTGGTGAAGAAATGTTGA 221  
 DB 40 Ile--ProAspIleAspSerAlaValArgIleIleProValAsnTyRAspSerAspPro 58  
 OY 222 AAGGAAGAGCTTATTTATTTAGTGAACCTTAAGTTCCTGAAATGTATCCAGGTA 281  
 DB 59 LysLeuAspSerGlnLeuTyThrValGluMetThrIleProAlaGlyValSerAlaVal 78  
 OY 282 ACGGCTTACCT-----GTTATGATGAAGATTAAGGCTTAGACGACTGTAAAT 332  
 DB 79. LysIleAlaProThrAspSerLeuThrSerSerGlyGlnIleGlyLeuValAsn 98  
 OY 333 ACCGCTATGCTTCCCATCAATATCTACAGATTGTGTGATGAAGAAAGAAAAATG 392  
 DB 99 ValAsnAsnProAspGlnAsnMetAsnTyRTrile-----GAGTTTACA 419  
 OY 393 TTTAAAGATCATGGTGCA-----GAGTTTACA 419  
 DB 111 ArgLysAspSerGlyAlaGlyAsnPheMetAlaGlyGlnLysGlySerPheProValLys 130  
 OY 420 CCTAATCAACAAATTAAGCTTTAAAGCGCTGAATTAAGTACTAGCGGGGAA-----AAA 470  
 DB 131 GluAenthrSerTyThrPheSerAlaIle---TyrThrGlyGlyGlyProAsnSer 149  
 OY 471 AAAATATCTCCTGGAATATATACATCAGCTTATGCTGTACTATGTAAC 524  
 DB 150 GlyTyrSerSerGlyThrTyraIaGlyAsnLeuThrValSerPheTySerAsn 167  
 RESULT 7  
 ADR14917  
 ID ADR14917 standard; protein; 146 AA.  
 XX  
 AC ADR14917;  
 DT  
 XX  
 XX 04-NOV-2004 (first entry)  
 XX  
 DE CS6 related protein #2.  
 XX  
 KM CS6; csaA, csaB, csaC, csaD, origin of replication; lac promoter;  
 KM kanamycin resistance gene; CS6A; CS6B; enterotoxigenic E. coli.  
 OS Unidentified.  
 XX  
 CC US2004156829-A1.  
 XX  
 PN 12-AUG-2004.  
 PD  
 PP 02-FEB-2004; 2004US-00768093.  
 XX  
 PR 13-MAY-1994; 94US-00243482.  
 PR 24-JAN-1997; 97US-00788145.  
 PR 10-JAN-2000; 2000US-00479877.  
 XX  
 PA (USSA ) US SEC OF ARMY.  
 PI Wolf MK, Casseels FJ, Boedeker EC;  
 PI WPI; 2004-592722/57.  
 DR  
 XX  
 PT Inducing the production of antibodies against CS6 protein, useful for  
 PT preventing pathological effects of enterotoxigenic E. coli, by  
 PT administering a composition comprising transformed bacteria producing CS6  
 PT antigens.  
 XX  
 PS Claim 6; SEQ ID NO 10; 22pp; English.  
 XX  
 CC The invention relates to a method of inducing, in a susceptible host, the  
 CC production of antibodies against a CS6 protein, comprising administering

CC a composition of matter comprising bacteria transformed with a plasmid  
 CC containing genes *csbA*, *csbB*, *csbC* and *csbD*, an origin of replication, a  
 CC *lac* promoter and a kanamycin resistance gene, where the bacteria  
 CC expresses both *CS6A* and *CS6B* proteins. The invention also relates to a  
 CC composition of matter comprising a protein in a pharmaceutical carrier.  
 CC The pharmaceutical carrier is a carbonated liquid. The method is useful  
 CC for inducing the production of antibodies against the *CS6* protein. The  
 CC protein and composition are used as vaccines for preventing pathological  
 CC effects of enterotoxigenic *E. coli*. This sequence represents a *CS6*  
 CC related protein of the invention.

XX Sequence 146 AA;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:
Score: 2,84e-05	146	43	22	55	31
Percent Similarity: 43.0%					
Best Local Similarity: 28.5%					
Query Match: 12.1%					
DB:	8	Gaps:	7		

US-10-768-093-4 (1-527) x ADR14917 (1-146)

QY 135 AAAAATCTCCAGTATCAAGCACTATTGCAAAAGTTTTCGACCTCGAA----- 185  
 DB 6 LYSERLEUAPVALAENVALAENILEGLUGLNAENPHILE--PROAPILLEAPSER 24  
 QY 186 ---CCACGCAATACAGCCTTCTTTGGTGAATAATGTTGAAGAGAGACCTTATATTT 242  
 DB 25 AIAVALAGLILIELEPROVALAENTYRABSERAPROVALAENLAPSERGINSLEUTYR 44  
 QY 243 AGTGTGAATTAACCTGTTCTTGAATAATGTTATCCAGTAAAGCTTCACTCT----- 293  
 DB 45 THRVALGLUMETHTHLEPROVALAENLVALAVALYILELAPROTHRAPSER 64  
 QY 294 GTTTATGATGAAGATTATGAGGACGACTAGTAAATACCGGTCGATCTTCCCAATCA 353  
 DB 65 LEUTRISERSEGLYGLNGLINILEGLYSLAENVALAENAPROAPGINSLE 84  
 QY 354 ATPATCTACCAAGATTGTTGATGAGAAAGGAAAAAATGTTAAAGATCATGTCGA--- 410  
 DB 85 METAENTYRTRYILE-----ARGYLAPSERGLYLAAGLY 96  
 QY 411 -----GAGGTACACCTTAATCAACAAATAACTTT 440  
 DB 97 AANPHEMELAAGLYGLNLYSGLYSERPHEPROVALYGLAENAPTHSERTRYRHPHE 116  
 QY 441 AAAAGCGTGAATTATTAACGCGGGA-----AAAAAATATCTCTGGAATATAT 491  
 DB 117 SERALALIE--TYRTHIGLYGLYGLYTYRPROBANSERGLYRSEISERGLYTHRYR 135  
 QY 492 AACGATCAGATTATGTTGTTACTATATTAAC 524  
 DB 136 AAGLYAENLEUTHVALSERPHERYSERAN 146  
 RESULT 8  
 ID AAE37981 standard; protein; 221 AA.  
 XX AAE37981,  
 AC AAE37981,  
 XX 06-NOV-2003 (first entry)  
 DT XX  
 DE Human kinase and phosphatase (KPP-26) protein.  
 XX  
 XX Human; kinase; phosphatase; KPP; cell proliferative disorder; hepatitis;  
 XX atherosclerosis; cirrhosis; haemoglobinuria; polycythaemia vera; cancer;  
 XX psoriasis; chromoblastoma; developmental disorder; Reiter's syndrome;  
 XX renal tubular acidosis; anaemia; mental retardation; Alzheimer's disease;  
 XX neurological disorder; Parkinson's disease; epilepsy; glomerulonephritis;  
 XX autoimmune disorder; inflammatory disorder; AIDS; Goodpasture's syndrome;  
 XX acquired immune deficiency syndrome; immunosuppressive; Crohn's disease;  
 XX noctropic; transgenic; dermatitis; multiple sclerosis; diabetes mellitus;

KW allergy; gout; Grave's disease; Hashimoto's thyroiditis; bowel syndrome;  
 KW osteoporosis; rheumatoid arthritis; Sjogren's syndrome; ophthalmological;  
 KW gene therapy; asthma; anticonvulsant; uterine; pancreatitis.  
 XX  
 XX Homo sapiens.  
 OS  
 PN W02003050084-A2.  
 PD 19-JUN-2003..

XX 19-JUN-2003..

PF 06-DEC-2002; 2002MO-US039126.

XX 07-DEC-2001; 2001US-0340235P.

PR 19-DEC-2001; 2001US-0343007P.

PR 21-DEC-2001; 2001US-0343546P.

PR 04-FEB-2002; 2002US-0354388P.

PR 15-FEB-2002; 2002US-0357675P.

XX (INCY-) INCYTE GENOMICS INC.

XX Kabir AE, Chien D, Wilson AD, Swarnakar A, Gorvad AE;  
 PI Hafalia ADA, Emerling BM, Ramkumar J, Jin P, Griffin JA, Margulis JP;  
 PI Baughn MR, Chawla NK, Lehr-Mason PM, Khare R, Lee S, Hawkins PR;  
 PI Becha SD, Lee SY, Sprague WW, Zebardjian Y;

DR WPI, 2003-532894/50.

DR N-PSDB; AAD57353.

XX New human kinases and phosphatases and polynucleotides, useful for  
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
 PT cancer or hepatitis.

XX Claim 1, Page 230; 282pp; English.

XX The invention relates to an isolated polypeptide, which is a human kinase  
 CC and phosphatase (KPP). KPP agonists and antagonists are useful for  
 CC diagnosing, treating or preventing disorders associated with aberrant  
 CC expression of KPP, particularly cell proliferative disorders (e.g.  
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal  
 CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary  
 CC chromocytoma or cancer), developmental disorders (e.g. renal tubular  
 CC acidosis, anaemia or mental retardation), neurological disorders (e.g.  
 CC Alzheimer's disease, Parkinson's disease or epilepsy), autoimmune/  
 CC inflammatory disorders (e.g. AIDS; acquired immune deficiency syndrome,  
 CC allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's  
 CC disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,  
 CC gout, Grave's disease, Hashimoto's thyroiditis, irritable bowel syndrome,  
 CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's  
 CC syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,  
 CC bacterial, fungal, parasitic, protozoan or helminthic infections. The KPP  
 CC is useful in assessing the effects of exogenous compounds on the  
 CC expression of nucleic acids and kinases and phosphatases. KPP gene is  
 CC useful in gene therapy and for creating transgenic animals to model human  
 CC disease. The present sequence is human KPP protein

XX Sequence 221 AA;

XX Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:
Score: 0.0198	221	41	27	72	21
Percent Similarity: 92.50					
Best Local Similarity: 42.2%					
Query Match: 25.5%					
DB:	6	Gaps:	7		

US-10-768-093-4 (1-527) x AAE37981 (1-221)

QY 93 CTGCTGATTCGCGGCAAGCAGCAAGAAATAGGACTTAAATACTCCAGTATCA 152  
 DB 41 MCTHTGTYRGLYSEHLSERHLSERVALTYRSEGLINSERLYASNNILEPROLEA 60  
 QY 153 -----ACAGCATTTGCAAAAGTTTTCGACCTGAGACAGCAATACAG--- 197

Accession	Protein	Length	Source	Organism	Accession	Protein	Length	Source	Organism
D0	GlnProAlaThrThrThrValSerThrSerLeuProValProAlaProSerLeuProTyr	80			D0	GlnProAlaThrThrThrValSerThrSerLeuProValProAlaProSerLeuProTyr	80		
QY	198 -----CCTCTCTTTGGTGAATAATTGGAAAGGAGAGCTTATT	239			QY	198 -----CCTCTCTTTGGTGAATAATTGGAAAGGAGAGCTTATT	239		
D0	81 GluGlnThrIleValAlaPheProGlySerThrGlyHisIleValValThrSerAlaSer	100			D0	81 GluGlnThrIleValAlaPheProGlySerThrGlyHisIleValValThrSerAlaSer	100		
QY	240 TTATGTGTGAAC-----TTAACTGTTCCTGAAATATGTA-----TCCCAAGTA	281			QY	240 TTATGTGTGAAC-----TTAACTGTTCCTGAAATATGTA-----TCCCAAGTA	281		
D0	101 ThrSerValIleThrGlyGlnValLeuGlyGlyProHisLeuHisLeuMetArgSerThrVal	120			D0	101 ThrSerValIleThrGlyGlnValLeuGlyGlyProHisLeuHisLeuMetArgSerThrVal	120		
QY	282 ACGGTCTAACCTCTTTATGATGATGAAGATTATGGCTTATGACGACTAGTAATAATCCGCTAT	341			QY	282 ACGGTCTAACCTCTTTATGATGATGAAGATTATGGCTTATGACGACTAGTAATAATCCGCTAT	341		
D0	121 SerIleLeuIleThrTyr--GlnYscYsGlyLeuLysArgSerGluGlnIleGlu	139			D0	121 SerIleLeuIleThrTyr--GlnYscYsGlyLeuLysArgSerGluGlnIleGlu	139		
QY	342 GCTTCGCCCAATGAATATATCTACACGATTTGCTGATGAGAAAGGAAAAAATGTAAAGAT	401			QY	342 GCTTCGCCCAATGAATATATCTACACGATTTGCTGATGAGAAAGGAAAAAATGTAAAGAT	401		
D0	140 AsnThrSerSerVal-----GlnIleIleGluGlnHis--ProProMetIleGlnAsn	156			D0	140 AsnThrSerSerVal-----GlnIleIleGluGlnHis--ProProMetIleGlnAsn	156		
QY	402 CATGCTGCAGAGGTTACACTTAATCAACAAATTACTTTTAAAGCGCTGATTAATACTAC	461			QY	402 CATGCTGCAGAGGTTACACTTAATCAACAAATTACTTTTAAAGCGCTGATTAATACTAC	461		
D0	157 AsnAlaSerGlyAlaThrValAlaThrAlaThrThrSerThrAlaThrSerLysAsnSer	176			D0	157 AsnAlaSerGlyAlaThrValAlaThrAlaThrThrSerThrAlaThrSerLysAsnSer	176		
QY	462 GGGGAAAAAAAATATCTCTCTGGAATATTAACGATCAGCTATAGTGTGTTACTATGTA	521			QY	462 GGGGAAAAAAAATATCTCTCTGGAATATTAACGATCAGCTATAGTGTGTTACTATGTA	521		
D0	177 GlySerProThrIleHisProSerGlnTyrProAlaGlnPheAlaHisGlnThrTyrIle	196			D0	177 GlySerProThrIleHisProSerGlnTyrProAlaGlnPheAlaHisGlnThrTyrIle	196		
QY	522 AAC 524				QY	522 AAC 524			
D0	197 Ser 197				D0	197 Ser 197			
RESULT 9					RESULT 9				
ADW47927	standard; protein; 268 AA.				ADW47927	standard; protein; 268 AA.			
XX	ADW47927;				XX	ADW47927;			
AC	ADW47927;				AC	ADW47927;			
DT	07-APR-2005 (first entry)				DT	07-APR-2005 (first entry)			
XX	Thermobifida fusca protein kinase C-terminal domain.				XX	Thermobifida fusca protein kinase C-terminal domain.			
DE	Protein kinase; antibacterial; tuberculostatic.				DE	Protein kinase; antibacterial; tuberculostatic.			
XX	Thermobifida fusca.				XX	Thermobifida fusca.			
XX	WO2005007880-A1.				XX	WO2005007880-A1.			
PN	27-JAN-2005.				PN	27-JAN-2005.			
PD	19-JUL-2004; 2004WO-1B003096.				PD	19-JUL-2004; 2004WO-1B003096.			
PF	18-JUL-2003; 2003US-0487943P.				PF	18-JUL-2003; 2003US-0487943P.			
XX	(INSP ) INST PASTEUR.				XX	(INSP ) INST PASTEUR.			
PA	Alzart P, Botel B, Villarino A, Fernandez P, Cole S;				PA	Alzart P, Botel B, Villarino A, Fernandez P, Cole S;			
PI	WPI; 2005-112885/12.				PI	WPI; 2005-112885/12.			
XX	Identifying substances that modulates pknB protein kinase or pspB2				XX	Identifying substances that modulates pknB protein kinase or pspB2			
PT	phosphatase activity, for identifying antibacterial substances, by				PT	phosphatase activity, for identifying antibacterial substances, by			
PT	comparing the activity from cells contacted with the substance to cells				PT	comparing the activity from cells contacted with the substance to cells			
PT	not contacted with the substance.				PT	not contacted with the substance.			
XX	Example; SEQ ID NO 6; 64bp; English.				XX	Example; SEQ ID NO 6; 64bp; English.			
PS	The present sequence is that of the C-terminal domain of a Ser/Thr				PS	The present sequence is that of the C-terminal domain of a Ser/Thr			
CC	protein kinase of Thermobifida fusca. Sequence homology is shown to the C				CC	protein kinase of Thermobifida fusca. Sequence homology is shown to the C			
CC	-terminal domain of Ser/Thr protein kinase PknB ADW47924 of Mycobacterium				CC	-terminal domain of Ser/Thr protein kinase PknB ADW47924 of Mycobacterium			
CC	tuberculosis M. tuberculosis PknB, PknA and protein phosph								

CC	or PeSP (or related sequences) . Such substances are potential
CC	antibacterial agents.
XX	Sequence 268 AA:
Alignment Scores:	
Pred. No.:	0.0396
Score:	90.50
Percent Similarity:	42.1%
Best local Similarity:	26.8%
Query Match:	9.6%
DB:	Gaps: 10
US-10-768-093-4 (1-527) x ADW47927 (1-268)	
QY	69 AAACATATGGTTTAAATCTTAAT-----CTTGCTCATTCGACGACCATGCC 116
Db	91 LysGIuValGIgLYLeuThrAspIleValGIuAspGIuIleThrSerPheAspAsnProPro 110
QY	117 AGACAGAAATAGCGACTAAAACCTTCCA-----GTA 149
Db	111 GlyThrValIleThrThrLysProAlaProGIuLysAlaAsnArgGIuLysSerVal 130
QY	150 TCACAGCATTTTCAAAAAGTTTTCGACCTGACACGACGATCAGCCTCTTTGGT 209
Db	131 ThrLeuThrIleSerAlaGlyPhe-----PrometProAsnVal-----ValGly 145
QY	210 GAATATGTGGAAAGAGAGAGCTTTATTTATTTAGTGAACCTTAACCTTCTCGAAAT 269
Db	146 GlnLysValAspAlaArgArgLeuLeuGlnSerSerAspLeu----- 160
QY	270 GTATCCAGGTAACGGTCTACCCCTGTTATGATGAAATATATGGTTAGACGACTAGTA 329
Db	161 -----GluValThrValValGIuLysIleHisAspGIuValProGIuLysIleValIle 178
QY	330 AAT-----ACCGCTATGCTCTCCCAATCATATA----- 356
Db	179 SerGIuLysProGIuLysIleValThrThrValGIuLysIleGlnSerValThrLeuThrVal 198
QY	357 -----ATCTACCAAGATTGTGAT-----GAGAAAGG 383
Db	199 SerSerGIyProGIuLysIleValGIuValProAspIleArgGIyTrpLysValAspLysAla 218
QY	384 AAAAATATGTTAAAGATCATAGTGGACAGAGGTATACCACTTAATCAACAATATCTTTATA 443
Db	219 ArgLysGIuLeuGIuLysArgLysPheIleValThrValHisGlnValIleLysAsnArg 238
QY	444 GCGCTGATTTACTT-----AGCGGGGAAAAAAAATATATCTCTGGAATATATACATCAG 500
Db	239 ValGIuLysArgLysProLysArgLysGlu-----AlaPro-LysGIySerThrIleGI 255
QY	501 GTTATGG 507
Db	255 uileTrp 257
RESULT 10	
ABR41065	
ID	ABR41065 standard; protein: 911 AA.
XX	ABR41065;
AC	
DT	22-MAY-2003 (first entry)
XX	
DE	Human MAP kinase cascade activator #50.
XX	
XX	Human, E1k1 phosphorylation; E1k1 phosphorylation kinase; virucide;
KW	antiinflammatory; immunomodulator; cytostatic; antiallergic; anti-HIV;
KW	antiinflammatory; antiarthritic; antidiabetic; anticancer; gene therapy;
KW	inflammation; autoimmune disease; viral disease; cancer; diabetes;
KW	rheumatoid arthritis; asthma; allergic rhinitis; AIDS; viral hepatitis;
KW	IGA nephritis.
XX	
XX	Homo sapiens
DS	



Dd		74	GIUGlInThrIleValPheProGlySerTrnGlyHisIleValValThlSerAlaSerSer	93
Oy		240	TTTAGTGTGAAC-----TTACTGTCCTGAAATATGA-----TCCCAAGTA	281
Dd		94	ThrSerValThrGlyGlnValLeuGlyGlyProHlaAenLueMetArgArgSerThrVal	111
Oy		282	ACGGTGTAACCTCTTTATGATGAAGATTATGGTTATGAGACGACTGATAATACCGCTCAT	341
Dd		114	SerLeuLenAphTrpYyr--GlnLySvGgLYLeuLySarGLysSerGluGlnIleGlu	133
Oy		342	GCTTCCCAATCAATTAATTAACAAGATTGTGATGAGAAAGGAAAAAATGTTAAAAGAT	401
Dd		133	AenTrhSerSerVal-----GlnIleIleGluGlnHis---ProprometIleelnasn	141
Oy		402	CATGCTGCAGAGGTTACACCTTAATCAACAATAACTTTTAAAGCCTGAAATTAATACG	461
Dd		150	AenAlaSerGlyAlaThrValAlaThrAlaThrAlaThrTrhPheThrAlaThrSerLyasnsSer	161
Oy		462	GGGGAAAAAAAAAATATCTCTCGAATATAT	491
Dd		170	GlySerAsn-----SerGluGlyAepTry	177
 RESULT 12				
AD	C95115	ID	ADC95115 standard; protein; 378 AA.	
XX		AC	ADC95115;	
DT		01-JAN-2004	(first entry)	
DE		E.	fascium protein sequence SEQ ID 4742,	
KM		Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;		
XX		abdominal-pelvic infection.		
OS		Enterococcus faecium.		
XX		US6583275-B1.		
PN		24-JUN-2003.		
PF		30-JUN-1998;	98US-00107532.	
PR		02-JUL-1997;	97US-0051571P.	
PR		14-MAY-1998;	98US-0085598P.	
PA		(GENO-) GENOME THERAPEUTICS CORP.		
PI		Doucette-Stamm LA, Bush D;		
PI		WPI; 2003-799836/75.		
DR		N-PSDB; ADC91461.		
PT		New isolated nucleic acid derived from Enterococcus faecium encoding an		
PT		Enterococcus faecium polypeptide useful for detection, prevention and		
PT		treatment of a pathological condition resulting from a bacterial		
PT		infection.		
XX		Example 1; SEQ ID NO 4742; 243bp; English.		
PS		The invention relates to an isolated nucleic acid derived from		
XX		Enterococcus faecium encoding an Enterococcus faecium polypeptide having		
CC		one of 10 fully defined sequences given in the (or comprising 40		
CC		sequential nucleotides chosen from any of the nucleic acids, its		
CC		complement or sequences hybridizing to it). Also included are a		
CC		recombinant vector comprising the nucleic acid operably linked to		
CC		a transcription regulatory element, a cell comprising the vector and a		
CC		single-stranded probe comprising the nucleic acid. The nucleic acids are		
CC		chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.		
CC		The nucleic acids is useful for diagnosing pathological conditions		
CC		resulting from E. faecium bacterial infection (e.g. urinary tract		
CC		infection), bacteraemia, endocarditis, wounds and abdominal-pelvic		
CC		infection) and for screening drugs such as agonists and antagonists. The		

CC	nucleic acid is useful for recombinant production of <i>Candida albicans</i> -	
CC	derived peptides or antisense polypeptides. Pharmaceutical compositions	
CC	and vaccines containing the nucleic acid are useful for preventing or	
CC	treating <i>Enterococcus faecium</i> infections. The present sequence represents	
CC	one if the disclosed <i>E. faecium</i> proteins.	
XX		
SO	Sequence 378 AA;	
	Alignment Scores:	
	Pred. No.:	0.191
	Score:	86.00
	Percent Similarity:	42.7%
	Best Local Similarity:	26.0%
	Query Match:	9.2%
		Indels: 30
		Gaps: 8
US-10-768-093-4 (1-527) x ADC95115 (1-378)		
QY	90 ATTTCTGCTTCATTCGGAGCCATGCGCAAGAAACAGAAATAGGACTAAAACTTCCACATA	149
DB	158 ValLeuAlaAlaAlaGlyAlaGlyThrGlnGlyAspTyrThrGlyThSerPhe-----	175
QY	150 TCAACGACTATTTGAAAAGTTTCTTTTTCACCTGACACAGATACAGCTTCCTTGGT	209
DB	176 ---ThSerIleGlyHisGlyAlaArgPheThrProGluAlaGlyAlaGlnProAlaIleGly	194
QY	210 GAATAATGTGGAAAG---GAAAGAGCTTATATTATAGTGTGAACCTTAACCTTCTGAA	266
DB	195 Lys---ValGlyLysThrGlnGlnValGlnGlnAlaLysValGlnValIleLeuProGln	213
QY	267 AATGTATCCCAAGGTAACGCTTAC-----	296
DB	214 ThrIleGlnLysGlnValIleGlnAlaMetArgSerAlaHisProTyrGlnGluProAla	233
QY	297 TATGAT-----	332
DB	234 TyrAspLeuPheAlaIleAspGlnProValGlnMetPheGlyLeuGlyArgVal-----	251
QY	333 ACCCGTAGTCTTCCCATTCATATATATACAGATTTGTGATGAGAAAAGGAAAAAATG	392
DB	252 ---GlyGlnLeuProGlnGlnIlePheIleGlnAlaPheValGlnGlnValLysGlnAla	270
QY	393 TTTAAAGATCATCTGGTCAGAGCTTACA-----CCTAATCAACAATAACTTTTAA---	443
DB	271 PheGlnLeuAspGlyLeuArgIleValGlnProLysAsnAlaLysSerSerValLysArg	290
QY	444----GGCGTGAATTAATCAAGCGGGGAAAAA 470	
DB	291 IleAlaIleCysGlyGlySerIleValLys 300	
RESULT 13		
AAAB49208		
ID	AAAB49208 standard; protein; 506 AA.	
XX		
AC	AAAB49208;	
XX		
DT	12-MAR-2001 (first entry)	
XX		
DE	V.cholerae VPI phage conserved gene ALA protein.	
XX		
KW	Bacteriophage; pathogenicity island; vaccine; allergy.	
XX		
OS	Unidentified.	
XX		
PN	WO200067784-A1.	
XX		
PD	16-NOV-2000.	
XX		
PF	10-MAY-2000; 2000WO-US012580.	
XX		
PR	10-MAY-1999; 99US-0133373P.	
XX		
PA	(UTMA-) UNIV MARYLAND BALTIMORE.	



XX Karoalle DK;  
 PI  
 DR WPI/ 2001-122703/13.  
 XX

XX Isolated bacteriophage encoded by a pathogenicity island of a pathogenic  
 PT bacterium, useful for generating bacteriophage-base vaccines or vectors,  
 XX e.g. to treat allergies.  
 PT

XX Disclosure, Fig 3, 59pp; English.  
 XX

XX The present invention relates to an isolated bacteriophage encoded by a  
 CC pathogenicity island of a pathogenic bacterium. The bacteriophage may be  
 CC used in pharmaceuticals e.g. for generating an immune response,  
 CC especially to produce antibodies, as a bacteriophage-based vaccine  
 CC against pathogenic bacteria expressing virulence factors, as a vaccine  
 CC vector, e.g. to prevent or treat allergies, diseases or other  
 CC pathological conditions  
 CC

XX Sequence 506 AA;  
 SQ

Alignment Scores:	0.279	Length:	506
Pred. No.:	85.00	Matches:	51
Score:	36.24	Conservative:	17
Percent Similarity:	27.1%	Mismatches:	62
Best Local Similarity:	9.1%	Indels:	58
Query Match:	4	Gaps:	10

US-10-768-093-4 (1-527) x AAB49208 (1-506)

QY 78 GGTATATCTATTTGCTTCTTCGCGAGCCAGTCCAGAAACAGAAATAGCGATTA 137  
 DB 211 GYVALLIIEAENVAlaValaEnglyPheGlySerGluAlaGlyAmaAlaLeuAlaThrSer 230  
 QY 138 AAC-----TTCCAGATCAAGACGATTTTCAAAAGTTT----- 176  
 DB 221 GlnAglIleAerlyleuAlaPheThrGlySerThGluIleGlyAenHisIleLeuLys 250  
 QY 177 GCACCTGAACACGATACAGCCTTTGTTGTAAGAAATGTTGGAAG----- 224  
 DB 251 GYVAlaIaAspAenleuIleProSerThrIleGluLeuGlyGlyLysSerProAenIle 270  
 QY 225 -----GAGAGAGCT 233  
 DB 271 TyrPheProAerIlePheSerHisGluAerGlnTyrLeuAerLysCysIleGluGlyAla 290  
 QY 234 TTAATATTAGTGTGAAC-----TTAAGTCTTCGAAATGATCCAGATA 281  
 DB 291 LeuLeuAlaPhePheAenGlnGlyValCysThrCysPro-----SerArgIle 307  
 QY 282 ACCGCTTAC--CCTGTTATGAT-----GAGATTTAGGTTA 317  
 DB 308 LeuValHisGluSerIleTyrGluLysPheIleAlaIleIleGluArgValAlaLeu 327  
 QY 318 GACACGACTGATAATACCGCTGCTTCCCAATCAATATCAACAGATGTTGATGAG 377  
 DB 328 IleLysGlnGlyAerProLeuAerThrGlnThrGlnIleGlyAlaGlnValSerLysGln 347  
 QY 378 AAAGGAAAAAATGTTA-----AAAGATCATGTCAGAGGTAA 419  
 DB 348 GlnTyrAerlyleuIleuGlyTyrIleGlnIleGlyLysAerProLysGlnValLeuIle 367  
 QY 420 -----CCTAATCAACAAATGATTTAAAGCGCTGATTAATACAGGGGGA 467  
 DB 368 PheGlyGlyHisAerProAenGln-----GluAenTyrLeuSerGlyGly 382  
 QY 468 AAAAAAATCTCTGCAATAT 491  
 DB 383 TyrTyrIleLysAerThrLeuPhe 390

RESULT 14  
 ABB49234

ID ABB49234 standard; protein; 541 AA.  
 XX  
 AC ABB49234;  
 XX  
 XX 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #34761.  
 XX

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW

XX Vibrio cholerae.  
 OS

XX MO20027183-A2.  
 PN

XX 03-OCT-2002.  
 PD

XX 21-MAR-2002; 2002WO-US009107.  
 PF

XX 21-MAR-2001; 2001US-00815242.  
 PR

XX 06-SEP-2001; 2001US-00948993.  
 PR

XX 25-OCT-2001; 2001US-0342923P.  
 PR

XX 08-FEB-2002; 2002US-00072851.  
 PR

XX 06-MAR-2002; 2002US-0362699P.  
 PR

XX (ELT-) ELITRA PHARM INC.  
 PA

XX Wang L, Zamudio C, Malone C, Hasselbeck R, Ohlsen KL, Zyskind JM,  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 PI

XX WPI/ 2003-029926/02.  
 DR

XX N-PSDB; ACA53104.  
 DR

XX Claim 25; SEQ ID NO 77158; 1766pp; English.  
 PS

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 623 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC the gene product or that has an activity against a biological pathway;  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising a strain in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_poc\_sequences  
 CC

XX Sequence 541 AA;  
 SQ

Alignment Scores:



Pred. No.:	0.288	Length:	54
Score:	85.00	Matches:	51
Percent Similarity:	36.2%	Conservative:	17
Best Local Similarity:	27.4%	Mismatches:	62
Query Match:	9.1%	Indels:	58
DB:	6	Gaps:	10

US-10-768-093-4 (1-527) X ABU49234 (1-541)

```

Oy 8GTTAAATTCTAAATCTTCTTCATCTCGGAGCGATCGGCAACAGAAATAGGAGCTAAA 137
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 G|yAa|l|eAenVAl|VAl|enG|yPheG|ySeG|uA|eA|yA|nAl|eA|uA|eA|A|n|Se 255
Oy 138 AAC-----TTCCAGATACAGCACTATTTCAAAAGATTGTTTT-- 176
Db G|nA|Tg|l|eA|e|P|y|e|u|e|A|A|Phe|n|Tg|y|Se|T|H|G|u|l|e|G|y|e|n|H|s|l|e|u|y|S 285
Oy 177 GCACTGAAACCAAGATACAGCTCTCTTTGGGAAATGTTGGAAAG----- 224
Db G|yA|A|A|A|A|eA|P|A|V|n|e|u|l|e|P|ro|Se|T|H|T|l|e|G|y|e|u|G|y|l|y|S|e|P|ro|A|e|n|l|e 305
Oy 286 C|yA|A|A|A|A|eA|P|A|V|n|e|u|l|e|P|ro|Se|T|H|T|l|e|G|y|e|u|G|y|l|y|S|e|P|ro|A|e|n|l|e 305
Oy 225 -----GAAAGAGCT 233
Db T|y|T|H|e|P|ro|A|e|n|l|e|Phe|Se|T|H|e|G|y|A|e|P|G|n|T|y|l|e|uA|e|P|y|e|S|y|l|e|G|u|G|y|A|A 325
Oy 234 TTATTATTAGTGTGTAAC-----TTAACTGTTCCTGAAATATGATCCAGTA 281
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
326 L|e|u|e|A|A|A|Phe|Phe|n|e|n|G|n|G|u|l|G|y|A|C|y|T|H|C|y|S|P|ro|-----SeA|Tg|l|e 342
Oy 282 ACGGCTAC---CTGTATTATGAT-----GAAGATTATGGTTA 317
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
343 L|e|u|V|A|H|e|G|y|Se|T|l|e|y|T|G|y|l|y|Phe|l|eA|y|A|l|e|y|l|e|G|y|A|G|V|A|A|A|e|u 362
Oy 318 GCAAGCACTAGTAATACCGCTAGCTCCCATCAATATATACACAGATGGTATGATGAG 377
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
363 I|e|y|V|e|n|G|n|G|y|A|e|P|ro|u|eA|e|P|H|n|G|u|T|H|G|n|l|e|G|y|A|A|G|n|A|l|e|S|e|T|y|e|G|u 382
Oy 378 AAAGGAAAAAATGTTA-----AAAGATCATGGTGCAAGGTTACA 418
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
383 G|n|T|y|A|e|P|y|l|e|e|u|G|y|T|y|l|e|G|n|l|e|G|y|l|y|A|e|P|G|n|G|u|l|A|G|u|l|e|n|l|e 402
Oy 420 -----CTTAATCAACAAATTAACCTTTAAAGCGGTGAATTAATACAGGGGGA 466
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 PheG|y|G|n|H|e|P|ro|A|e|n|G|n|-----G|u|S|e|n|T|y|l|e|u|S|e|T|y|G|y 417
Oy 468 AAAAATAATCTCTGCAATATAT 491
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
418 T|y|T|y|T|l|e|y|e|P|ro|T|H|C|y|e|P|H|e 425

```

RESULT 15  
ADR86415  
ID ADR86415 standard; protein; 1239 AA.

PA (ELIT-) ELITRA PHARM INC.  
PA (ELIT-) ELITRA CANADA LTD.  
XX  
XX  
XX Jiang B, Hu W, Lemieux S, Roemer T;  
PI  
DR MPI; 2004-594200/57.  
DR N-PSDB; ADDR5828.  
XX  
XX  
XX New purified or isolated *Aspergillus fumigatus* nucleic acid molecule  
PT encoding a gene product, useful for diagnosing and/or treating invasive  
PT fungal infections, such as Farmer's lung disease.  
XX  
XX Claim 1; SEQ ID NO 3465; 164pp; English.  
PS

The present invention relates to *Aspergillus fumigatus* genes that are essential and are potential targets for drug screening. The methods and compositions of the present invention are useful for diagnosing and/or treating invasive *Aspergillus fumigatus* infection, including the allergic forms of the disease, such as Farmer's lung disease. They can also be used in various drug discovery purposes, such as expression of the recombinant protein, hybridization assays and construction of nucleic acid arrays. The present sequence represents an *Aspergillus fumigatus* essential gene protein sequence, used during diagnosis and drug development in the invention. These genes share a high degree of sequence conservation with known essential genes of *Candida albicans*. The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format from WIPO.

Alignment Scores:	
Pred. No.:	0.488
Score:	84.50
Percent Similarity:	44.0%
Best Local Similarity:	28.0%
Query Match:	9.0%
DB:	8
Length:	1238
Matches:	35
Conservative:	20
Mismatches:	53
Indels:	17
Gaps:	6

US-10-768-093-4 (1-527) x ADR86415 (1-1239)

[illegible]

Search completed: April 10, 2006, 10:00:10  
Job time : 140.5 secs



QY 243 AGGTGAACTTAAGTCTTCTGAAAAATGATCCAGGTAACGGTCTTACCTGTTATGAT 302  
DB 61 SerValSerLeuThrValProGluValSerGlnValThrValTyrProValTyrAsp 80  
QY 303 GAAGATTATGGGTTAGAGCACTAGTAATACCGCTGATGCTTCCCAATCAATATCTAC 362  
DB 81 GluAspTyrGlyLeuGlyValLeuValLeuThrAlaAspSerGlnSerLeuLeuTyr 100  
QY 363 CAGATTGTTGATGAGAAAGGAAAAATGTTAAAAAGATCATGTCAGAGGTTACACCT 422  
DB 101 GlnIleValAspAspLeuGlyValArglySerLeuValAspHisGlyAlaGluValThrPro 120  
QY 423 AATCAACAATTAATCTTTAAACCGCTGAATTAATCTACTAGCGGGAATAAAAAATATCTCT 482  
DB 121 AengGlnIleThrPheArgAlaLeuAsnTyrThrSerGlyAspLeuGluLeuProPro 140  
QY 483 GGAATATTAAGATCAGATGCTTGGTGTGCTACTATGTAAC 524  
DB 141 GlyIleTyrAsnAspGlnValMetValGlyTyrTyrValAsn 154

## RESULT 2

183348  
CS6 structural subunit B - Escherichia coli  
C/Species: Escherichia coli  
C/Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 09-Jul-2004  
C/Accession: 183348  
R.Milshaw, G.A.; Smith, H.R.; McConnell, M.M.; Rowe, B.  
FEMS Microbiol. Lett. 49, 473-478, 1988  
A>Title: Cloning of genes encoding coli-surface (CS) antigens in enterotoxigenic Escheri-  
A/Reference number: 160266  
A/Accession: 183348  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-167 <RES-  
A/Cross-references: UNIPROT:P53510; UNIPARC:UPI000012A402; EMBL:U04844; NID:G442375; PID  
US-10-768-093-4 (1-527) x 183348 (1-167)

## Alignment Scores:

Pred. No.: 5,99e-05 Length: 167  
Score: 126.00 Matches: 49  
Percent Similarity: 45.8% Conservative: 28  
Best Local Similarity: 28.2% Mismatches: 75  
Query Match: 13.4% Indels: 16  
Gaps: 8

QY 63 ATGAAGAAACAAAT--CGTTAATCTTAATCTTCTGTCATGGGACCGATCGACGA 119  
DB 2 LeuLysLeuIleLeuProAlaIleValLeuIleAlaGlyThrSerGlyValValAsnAla 21  
QY 120 ACAGAAATAGCACTAAACCTTCCCATATCAACGATATTTCAAAAGTTTGTGCA 179  
DB 22 GlyAsnTrpGlnTyrLysSerLeuMetValAsnValAsnIleGluGlnAsnHelle--- 40  
QY 180 CCTGA-----CCAGATACAGCCTCTCTTGGGAAATGTTGAAAGCA 227  
DB 41 ProAspIleAspSerAlaValAlaArgIleIleProValAsnTyrAspSerProLysLeu 60  
QY 228 GGAAGCTTATTTAGTGTGAATTAATCTTCCGAAATGTAATCCAGGTAACGGTC 287  
DB 61 AsnSerGlnLeuTyrThrValGluMetThrIleProAlaGlyValSerAlaValLysIle 80  
QY 288 TACCT-----GTTATGATGAAAGATTATGGGTTAGAGCACTAGTAATACCGCT 338  
DB 81 ValProThrAspSerLeuThrSerGlyGlnGlnIleGlyLysLeuValAsnValAsn 100  
QY 339 GATGCTTCCCAATCAATTAATCTACAGATTTGTTGATGAGAA-----CGGAAAAAATG 392  
DB 101 AsnProAspGlnAsnMetAsnTyrTyrIleArgLysAspSerGlyAlaGlyLysPheMet 120  
QY 393 TTTAAAGATCATGCGCA---GAGGTTACACCTTAATCAACAATATCTTTAAAGCGTC 449  
DB 121 AlaGlyGlnLysGlySerPheSerValLysGluAsnThrSerTyrThrPheSerAlaIle 140

QY 450 AATTATCTACGGGGAA-----AAAAAATATCTCTGAAATATATACATCAG 500  
DB 141 ---TyrThrGlyGlyGlyTyrProAsnSerGlyTyrSerSerGlyThrTyrAlaGlyHis 159  
QY 501 GTTATGTTGGTCTACTATGTAAC 524  
DB 160 LeuThrValSerPheTyrSerAsn 167

## RESULT 3

T47930  
hypothetical protein T20K12.280 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C/Accession: T47930  
R.De Haan, M.; Maarse, A.C.; Grievell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quet  
submitted to the Protein Sequence Database, January 2000  
A/Reference number: 224480  
A/Accession: T47930  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-718 <DEH>  
A/Cross-references: UNIPROT:Q9M2C6; UNIPARC:UPI000009FPC8; EMBL:AL137898  
A/Experimental source: cultivar Columbia; BAC clone T20K12  
C/Genetics:  
A/Map position: 3  
A/Introns: 47/1; 408/1; 448/3; 512/1  
A/Note: T20K12.280  
C/Superfamily: Arabidopsis thaliana hypothetical protein T20K12.280

## Alignment Scores:

Pred. No.: 1.37 Length: 718  
Score: 85.50 Matches: 37  
Percent Similarity: 41.9% Conservative: 28  
Best Local Similarity: 23.9% Mismatches: 73  
Query Match: 9.1% Indels: 17  
Gaps: 2

US-10-768-093-4 (1-527) x T47930 (1-718)

QY 69 AAACAATGGTTAATCTTAATCTTCTGTCATGGGACCGATCGACGAAGAAATA 128  
DB 334 ArgIleLeuGlyLysIleLeuSerLeuProGluThrCysSerProAlaAspSerProArg 353  
QY 129 GCGACTAAACCTTCCCATATCAACGATATTTCAAAAGTTTGTGCACTGACCA 188  
DB 354 LeuIleProAlaHisAspLeuValSerThrLeuSerGlnThrGluGlnProGluIle 373  
QY 189 CGAATACAGCCTTCT-----TTGGGAAATGTTGGAAG 224  
DB 374 LeuGlnThrProGluThrSerSerAlaThrAsnAspLeuIleAspGluAspSerAspLys 393  
QY 225 GAAGAGCTTATTTAGTGTGAATTAATCTTCCGAAATGTAATCCAGGTAAC 284  
DB 394 AspAspAspThrLeuPheThrIleAspValSerValProAspAspTyrGlyAsnGluThr 413  
QY 285 GTTACCTGTTTATGATGAAGATTATGGGTTAGAGCACTAGTAATACCGCTATGCT 344  
DB 414 GluAsnIleAspAsnAspGluGluSerGluLeuAspProLeuSerGluThrCysSerSer 433  
QY 345 TCCCAATCAATTAATCTACAGATTTGTTGATGAGAAAGGAAAAAATGTTAAAGATCAT 404  
DB 434 SerValSerArgGluValGluAsnValAspGluAspValGluLysGluMetLeuAsnHis 453  
QY 405 GGTGCAAG-----GTTACACCTTAATCAACAATTAATCT 437  
DB 454 GlnAlaHisSerProAlaSerProLeuGluSerProThrThrCysSerValArgLeuThr 473  
QY 438 TTTAAAGCGCTGAATTAATCTACAGCGGAAAAAATATCTCTCT 482  
DB 474 GluCysLysGluThrAlaThrAspAlaGlnGlyLysLeuSerPro 488

## RESULT 4

T09437  
Probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) alda [similarity] - Vibrio cholerae  
C/Species: Vibrio cholerae  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C/Accession: T09437  
R/Karaoili, D.K.R.; Johnson, J.A.; Bailey, C.C.; Boedeker, E.C.; Kaper, J.B.; Reeves, P.  
Proc. Natl. Acad. Sci. U.S.A. 95, 3134-3139, 1998  
A/Title: A Vibrio cholerae pathogenicity island associated with epidemic and pandemic st  
A/Reference number: 216572; MUID:98169509; PMID:9501228  
A/Accession: T09437  
A/Structure: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-506 <KAP>  
A/Cross-references: UNIPROT:P23240; UNIPARC:UPI0000000A06; EMBL:AF034434; NID:G3004923;  
A/Experimental source: strain N16961  
C/Genetics:  
A/Gene: alda  
A/Function:  
A/Note: part of the pathogenicity island (VPI); associated with epidemic and pandemic st  
A/Description: catalyzes the oxidation of aldehydes  
C/Superfamily: NAD-dependent aldehyde dehydrogenase; aldehyde dehydrogenase homology  
C/Keywords: NAD; oxidoreductase

Alignment Scores:  
Pred. No.: 1.45 Length: 506  
Score: 85.00 Matches: 51  
Percent Similarity: 36.2% Conservative: 17  
Best Local Similarity: 27.1% Mismatches: 62  
Query Match: 9.1% Indels: 58  
DB: 2 Gaps: 10

US-10-768-093-4 (1-527) x T09437 (1-506)

Oy 78 GGTAAATTCATTCCTGCTTCATTCGGCAGCCATGCCAGAACAGAAATAGCGACTAA 137  
Db 211 GYVALLIENAVAIValaenGlypHeGlyserGluAaGlyAsnAlaLeuAlaTherSer 230  
Oy 138 AAC-----TTCCAGATCAACGACTATTTCAAAAAGTTTTTT-- 176  
Db 231 GlnrgIleArpLyLeuAlaPheThrGlySerThrgluIleGlyAsnHleleuLys 250  
Oy 177 GCACCTGAACACGAAATACAGCCTCTTTGGTGAATGTTGGAAG----- 224  
Db 251 CysAlaAlaPheAsnLeuIlePheSerThrlleGluLeuGlyLysSerProAsnIle 270  
Oy 225 -----GAAGAGCT 233  
Db 271 TyrPheProArpIlePheSerHleGluArpGlnTyLeuArpLysCysIleGluAla 290  
Oy 234 TTATTAATTCAGTGAAC-----TTAACTGTTCTGAAATGTAATCCAGGTA 281  
Db 291 LeuLeuAlaPhePheAsnGlnGlyValCysThrCysPro-----SerArgIle 307  
Oy 282 ACGGCTAC--CCTGTTTATGAT-----GAAGATTATGGGTTA 317  
Db 308 LeuValHleGluSerIleTyrgluLysPheIleAlaLysIleIleGluArgValAlaLeu 327  
Oy 318 GGAGCAGTAAATTAACCGCTGATGCTCCCAATCAATATATACAGATTGTGATGAG 377  
Db 328 IleLysGlnGlyAsnProLeuArpThrgluThrglnIleGlyAlaGlnValSerIysGlu 347  
Oy 378 AAAGGAAAAAATGTTA-----AAAGATCATGTCGACAGGTTACA 419  
Db 348 GlnTyArpLysIleLeuGlyTyrlleGlnIleGlyLysArpGlnGlyAlaGluLeuIle 367  
Oy 420 -----CCTAATCAACAATTAACCTTTAAAGCGCTGAATTATACAGCGGGA 467  
Db 368 PheGlyGlnHlePheProAsnGln-----GluAsnTyLeuSerIyGly 382  
Oy 468 AAAAAATATCTCCTGGAATATAT 491  
Db 383 TyrTyrlleLysProThrLeuPhe 390

RESUR 5  
A82276  
aldehyde dehydrogenase VC0819 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C/Species: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Mar-2004  
C/Accession: A82276  
R/Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, M.D.; Vamathavan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: A82035; MUID:20406833; PMID:10952301  
A/Accession: A82276  
A/Structure: preliminary  
A/Molecule type: DNA  
A/Residues: 1-541 <HEI>  
A/Cross-references: UNIPARC:UPI0000164B65; GB:AE004167; GB:AE003852; NID:99655268; PIRN:  
A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
C/Genetics:  
A/Gene: VC0819  
A/Map position: 1  
C/Superfamily: NAD-dependent aldehyde dehydrogenase; aldehyde dehydrogenase homology

Alignment Scores:  
Pred. No.: 1.47 Length: 541  
Score: 85.00 Matches: 51  
Percent Similarity: 36.2% Conservative: 17  
Best Local Similarity: 27.1% Mismatches: 62  
Query Match: 9.1% Indels: 58  
DB: 2 Gaps: 10

US-10-768-093-4 (1-527) x A82276 (1-541)

Oy 78 GGTAAATTCATTCCTGCTTCATTCGGCAGCCATGCCAGAACAGAAATAGCGACTAA 137  
Db 246 GYVALLIENAVAIValaenGlypHeGlyserGluAaGlyAsnAlaLeuAlaTherSer 265  
Oy 138 AAC-----TTCCAGATCAACGACTATTTCAAAAAGTTTTTT-- 176  
Db 266 GlnrgIleArpLyLeuAlaPheThrGlySerThrgluIleGlyAsnHleleuLys 285  
Oy 177 GCACCTGAACACGAAATACAGCCTCTTTGGTGAATGTTGGAAG----- 224  
Db 286 CysAlaAlaPheAsnLeuIlePheSerThrlleGluLeuGlyLysSerProAsnIle 305  
Oy 225 -----GAAGAGCT 233  
Db 306 TyrPheProArpIlePheSerHleGluArpGlnTyLeuArpLysCysIleGluAla 325  
Oy 234 TTATTAATTCAGTGAAC-----TTAACTGTTCTGAAATGTAATCCAGGTA 281  
Db 326 LeuLeuAlaPhePheAsnGlnGlyValCysThrCysPro-----SerArgIle 342  
Oy 282 ACGGCTAC--CCTGTTTATGAT-----GAAGATTATGGGTTA 317  
Db 343 LeuValHleGluSerIleTyrgluLysPheIleAlaLysIleIleGluArgValAlaLeu 362  
Oy 318 GGAGCAGTAAATTAACCGCTGATGCTCCCAATCAATATATACAGATTGTGATGAG 377  
Db 363 IleLysGlnGlyAsnProLeuArpThrgluThrglnIleGlyAlaGlnValSerIysGlu 382  
Oy 378 AAAGGAAAAAATGTTA-----AAAGATCATGTCGACAGGTTACA 419  
Db 383 GlnTyArpLysIleLeuGlyTyrlleGlnIleGlyLysArpGlnGlyAlaGluLeuIle 402  
Oy 420 -----CCTAATCAACAATTAACCTTTAAAGCGCTGAATTATACAGCGGGA 467  
Db 403 PheGlyGlnHlePheProAsnGln-----GluAsnTyLeuSerIyGly 417  
Oy 468 AAAAAATATCTCCTGGAATATAT 491  
Db 418 TyrTyrlleLysProThrLeuPhe 425

J04036  
 phosphohydroxyurate hydratase (EC 4.2.1.11) - fission yeast (Schizosaccharomyces pombe)  
 N/Alternate names: 2-phospho-D-glycerate hydrolyase; enolase  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 09-Jul-2004  
 C/Acession: J04036  
 R/Jackson, J.C.; Lopez, J.M.  
 Gene 154, 109-113, 1995  
 A/Title: A cDNA from Schizosaccharomyces pombe encoding a putative enolase.  
 A/Accession number: J04036, MUID:95172389, PMID:7867936  
 A/Aceesion: J04036  
 A/Molecule type: mRNA  
 A/Residues: 1-439 <JAC>  
 A/Cross-references: UNIPROT:P40370; UNIPARC:UPI000168FC9; GB:U13799; NID:9535441; PIDN:  
 C/Note: The authors translated the codon GTC for residue 20 as B and GAC for residue 83  
 C/Comment: This glycolytic enzyme catalyzes the dehydration of 2-phosphoglycerate to ph  
 C/Genetics:  
 A/Gene: enol  
 C/Superfamily: enolase  
 C/Keywords: carbon-oxygen lyase; gluconeogenesis; glycolysis; hydro-lyase

Alignment Scores:  
 Pred. No.: 1.8 Length: 439  
 Score: 84.00 Matches: 43  
 Percent Similarity: 40.44 Conservative: 24  
 Best Local Similarity: 25.94 Mismatches: 65  
 Query Match: 8.94 Indels: 34  
 DB: 2 Gaps: 9

US-10-768-093-4 (1-527) x J04036 (1-439)  
 QY 54 TGGTGTATATGAGAAAACATTGGT-----TTAATCTAATTCCTGCTCA 101  
 Db 131 TTPYTSYTRILEALAGLAAENPHEGLYTHRYSGLYPROTYRVALLEUPROVALPROSER 150  
 QY 102 TTC-----GGCAGCCATGCCAGAACAAATAAGGACGCTAAATACTGCCA 146  
 Db 151 PHEANVALLEUAANGLYGLYSERHISALAGLYLAAPLEUALAEPHEGLINGLUPHEMET 170  
 QY 147 GTATCAACGACTATTTCAAAAAGTTT----- 173  
 Db 171 ILELEUPROTHIRGVALAPROSERPHESEERGLUAMETAGTYPGLYALAGLINTHYR 190  
 QY 174 -----TTTGACCTGMAACCAACGAAATACAGCCTCTTTGGTGAAATGTGGGA 221  
 Db 191 GLINTHREULYSERILEALALEYLSERHISALAGLYSERSERALAGLY--AENVALGLY 209  
 QY 222 AAGGAGAGGCTTATTTAGTGTGAACCTTAACCTGTCCTGMAAATGTATCCAGGTA 281  
 Db 210 AEPGLUGLYGLYLE-----ALAPROASPLEUGINTHPRGGLINGLUALALEUASPLEU 227  
 QY 282 ACGGCTACCCCTGTTTATGATGATGAAGATTATGGGTGAGCAGCACTAGTAATACCGCTGAT 341  
 Db 228 ILEVALGLUALALEAENLYALAGLYTYRGU--GLYLSILELYALIEGLYLEUASBP 246  
 QY 342 GCTTCCCATCATATATCTACCAAGTTGTGATGAGAAAGGAAAAAATGTTAAAGAT 401  
 Db 247 VALTHRSERISERGLUPHELYR-----VALASP-----GLYLSYTYRASPLEUASPLE 262  
 QY 402 CATGGTGAAGAGGTATACCTTAATCAACAAATAAATCTTTAAAGCGCTGAAT-----TAT 455  
 Db 263 LYALALALALYSPROLYSPROGLUASNLYSLEUTHIRYRGINGLINTHTRASPLEUTYR 282  
 QY 456 ACTAGCGGGGAAAAAAA 473  
 Db 283 VALGLULEUSERLYEYS 288

RESULT 7  
 T39737  
 enolase - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: J39737  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z21875  
A:Accession: J39737  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-439 <MO>  
A:Cross-references: UNIPROT:P40370; UNIPARC:UPI0000129F3E; EMBL:AL050301; PIDN:CAB43486.  
A:Experimental source: strain 972h-; cosmid c1815  
A:Gene: SPDB:SPBC1815.01  
A:Map position: 2  
A:Superfamily: emolase

Alignment Scores:

Pred. No.:	1.8	Length:	439
Score:	84.00	Matches:	43
Percent Similarity:	40.4%	Conservative:	24
Best Local Similarity:	25.9%	Mismatches:	65
Query Match:	8.9%	Indels:	34
DB:	2	Gaps:	9

US-10-768-093-4 (1-527) x T39737 (1-439)

OY 54 TGGTGTATATGAGAAACAATTGGT-----TTAATTCTAATCTTGCTCA 101  
DB 131 Trrpstrlrlaenphaglythrlyaglyprotyrvalproser 150  
OY 102 TTC-----GGCAGCCATGCCAGAAAGAAATACGACTAATAAATCCCA 146  
DB 151 PheasnValleuabnglylserhlsalaglylaPheudalaphenglnuphemec 170  
OY 147 GTATCAACGACTATTTCAAAAGTTT----- 173  
DB 171 ltleuProthrghyalaproserPheserGualamerargtrpghyalagluThrtyr 190  
OY 174 -----TTGCACCTGAAACACGAAATACAGCTTCTTTGGTAAATGTTGA 221  
DB 191 Hiehrleuylserllealalybysargtyglyseraserlagnly--AanValgly 209  
OY 222 AAGGAGAGACTTATTATTATTAGTGACACTTAAGTTCCTGAAAATGTATCCAGTA 281  
DB 210 Asplolnlyglyle-----Alapromapleuclnhrproglngualaleuapleu 227  
OY 282 ACGGCTTACCTGTTTATGATGAAAGATTATGCGTTAGAGACGACTAGTAATACCGCTGAT 341  
DB 228 lleValglualalleuauylsalaglytyrglu---Glylysrleuylleglyleuasp 246  
OY 342 GCTTCCCAATCAATATCTACACAGATTGTTGATGAGAAAGAAAATAATGTAAAGAAT 401  
DB 247 ValAlaserbergluphetyr-----Valasp-----Glylystrapleuaple 262  
OY 402 CATGTCACAGAGGTTCACCTTAATCAACAATACTTTAAAGCGCTGAAT-----TAT 455  
DB 263 LysAlaAlaAlasprPolysProgluasnlyleuThrtyrGlnlnleuThrAspleuTy 282  
OY 456 ACTAGCGGGGAAAAAAA 473  
DB 283 ValgluLeuSerlylsys 288

RESULT 8  
JC6010  
formate C-acetyltransferase (EC 2.3.1.54) - Clostridium pasteurianum  
N:Alternate names: Pfl protein; pyruvate formate-lyase  
C:Species: Clostridium pasteurianum  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: JC6010; PC6004  
R:Weidner, G.; Sawers, G.  
J. Bacteriol. 178, 2440-2444, 1996  
A:Title: Molecular characterization of the genes encoding pyruvate formate-lyase and its  
A:Reference number: JC6010; MUID:96218720; PMID:8636053

A/Molecule type: DNA  
 A/Residues: 1-740 <ME11>  
 A/Cross-references: UNIPROT:Q46266; UNIPARC:UPI0000131758; EMBL:X93463; NID:g1072360; PI  
 A/Accession: PC6004  
 A/Molecule type: protein  
 A/Residues: 1-740 <ME12>  
 A/Cross-references: UNIPARC:UPI0000131758  
 C/Comment: This enzyme must be activated by the lyase-specific pyruvate formate-lyase-ac  
 ite and causes peptide cleavage.  
 C/Genetics:  
 A/Gene: pfl  
 A/Start codon: TTG  
 C/Complex: homodimer  
 C/Function:  
 A/Description: catalyzes the reversible conversion of acetyl-CoA and formate into pyruv  
 A/Pathway: anaerobic glucose metabolism  
 C/Superfamily: formate C-acetyltransferase 1; glycyl radical homology  
 C/Keywords: acyltransferase; coenzyme A; homodimer; lyase; thiolester bond  
 F/681-740/Domain: glycyl radical homology <GPR>  
 F/405/Active site: Cys (cysteine thiol radical intermediate) #status predicted  
 F/406/Active site: Cys (S-acetylcysteine intermediate) #status predicted  
 F/715/Active site: Gly (stable glycyl radical) #status predicted

Alignment Scores:  
 Pred. No.: 1.98 Length: 740  
 Score: 84.00 Matches: 39  
 Percent Similarity: 38.64 Conservative: 25  
 Best Local Similarity: 23.54 Mismatches: 68  
 Query Match: 8.94 Indels: 34  
 DB: 1 Gaps: 7

US-10-768-093-4 (1-527) x JC6010 (1-740)  
 Oy 93 CTTGCTTCATTCGGAGCGCATCCGACAGAAATAGGACTAAACCTCCAGTATCA 152  
 Db 326 TLeaEgLYa1GlytLeaPpGlyArGserLeuValThrLyAsnSerPheArgTyrLeu 345  
 Oy 153 ACGACTATTTCAAAAGTTTTCACCTGAAACGACGATACAGCTCTTTGGTGA 212  
 Db 346 HIsrHrleuLeuLeuLeuGlySerLArProGluProAmMetThrValLeuTrpSerL 365  
 Oy 213 AATGTTGAAAGAAAGAGCTTTATTATTAGTGGAACCTTAACCTGCTGAAATGTA 272  
 Db 366 AsnLeu---ProGluSerPheLysPheCysAlaGluMetSerLLeuThrAspSer 384  
 Oy 273 TCCAGAGTA-----ACGCTACCCCTGTTATGATGATGATGATGAGTTA 317  
 Db 385 TLeGlnTyrGluAsnAspPLeuMetArGProLLeuTyrGlyAspAspTyrAlaLea 404  
 Oy 318 -----GAGCAGACTAGTAAT 332  
 Db 405 CysCysValSerAlaMetArgValGlyLysAspMetGlnPhePheGlyAlaArgCysAsn 424  
 Oy 333 ACCGCTGATGCTTCCCAATCAATAATCTACAGATGTTGATGAGAAAGGAAAAATG 392  
 Db 425 LeuAlaLysCysLeuLeuLeuAlaLeuLeuGlyValAspGluLys----- 440  
 Oy 393 TTTAAAGATCATTGCGCAGAGGTTACACCTAAT---CAACAAATTACTTTAAAGCGCTG 449  
 Db 441 -----LysGlyTyrLeuValAlaProAspLLeuGluProLLeuThrAspLValLeu 457  
 Oy 450 AATTACTACGGGAGAAAAAATATCTCTCGAATATATACGAT---CAGTTATG 506  
 Db 458 AspTyr-----GluLysValLysGluAsnTyrPheLysValLeuGluLysMet 473  
 Oy 507 GTTGTTACTATGTAAC 524  
 Db 474 ALegLYLeuTyrValAsn 479

RESULT 9  
 JC7604  
 CD86 spliced variant CD86 deltaTM isoform - human  
 C/Species: Homo sapiens (man)

C/Date: 30-Jun-2001 #sequence revision 30-Jun-2001 #text\_change 18-Nov-2002  
 C/Accession: JC7604  
 R/Magistralli, G.; Caron, G.; Gauchat, J.F.; Jeanmin, P.; Bonnefoy, J.Y.; Delneste, Y.  
 Biochem. Biophys. Res. Commun. 280, 1211-1215, 2001  
 A/Title: Identification of an alternatively spliced variant of human CD86 mRNA.  
 A/Reference number: JC7604; MUID:21092744; PMID:11162656  
 A/Accession: JC7604  
 A/Molecule type: mRNA  
 A/Residues: 1-275 <MAG>  
 A/Cross-references: UNIPARC:UPI0000178C43  
 C/Comment: This CD86 variant expressed by activated human monocytes, is a costimulatory  
 C/Genetics:  
 A/Gene: cd8deltaTM  
 C/Superfamily: B-lymphocyte restricted antigen B7  
 C/Keywords: immune response

Alignment Scores:  
 Pred. No.: 1.86 Length: 275  
 Score: 83.50 Matches: 39  
 Percent Similarity: 39.8% Conservative: 29  
 Best Local Similarity: 22.8% Mismatches: 58  
 Query Match: 8.9% Indels: 45  
 DB: 2 Gaps: 6

US-10-768-093-4 (1-527) x JC7604 (1-275)  
 Oy 66 AAGAAACAATTGCTTAAATCTAAT-----CTTGCT 98  
 Db 109 LysLysProThrGlyMetLLeuGlyLHisGlnMetAsnSerGluLeuSerValLeuAla 128  
 Oy 99 TCATTGGCAGCCATGCCAGAACAGAAATAGCG-----ACTAAAACTTCCAGTATCA 152  
 Db 129 AsnHeserGlnProGluLLeuValProLLeuSerAsnLLeuThrGluAsnValTyrLeuAsn 148  
 Oy 153 ACGACTATTTCAAAAGTTTTCACCTGAAACGACGATACGAA----- 191  
 Db 149 LeuThrCysSerSerLLeuGlyTyrProGluProLysLysMetSerValLeuLeuArg 168  
 Oy 192 -----ATACGCTTCTTTGTTGTAATGTTGGA 221  
 Db 169 ThrLysAsnSerThrLLeuTyrAspGlyLLeuMetGlnLysSerGlnAspAsnValThr 188  
 Oy 222 AAGAGAGCTTATTATTATTATTAGTGGAACCTTAACCTGCTGAAATGTAATCCAGTA 281  
 Db 189 GluLeuTyrAspValSerLLeuSerLeuSerValSerPheProAspValThrSerAsnMet 208  
 Oy 282 ACGCTACCCCTGTTATGATGATGAT-----TATGGGTTA 317  
 Db 209 ThrLLeuPheCysLLeuGluThrAspLysThrArgLeuLeuSerSerProPheSerLLe 228  
 Oy 318 GAGCAGACTAGTAATACCGCTGATGCTTCCCAATCAATAATCTACAGATGTTGATGAG 377  
 Db 229 Gly-----ThrAsnThrMetGluArgGluGluSer-----Glu 239  
 Oy 378 AAGAGAAAAAATGTTAAAGATCATGTCAGAGGTTACACTTAACCAATAACT 437  
 Db 240 GlnThrLysLysArgGluLysLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuLLeuL 259  
 Oy 438 TTTAAAGCGCTGAATTTACTACGCGGGAATAA 470  
 Db 260 PheLysSerSerLysThrSerSerCysAspLys 270

RESULT 10  
 AF1226  
 hypocholesterol protein lmo1214 [imported] - listeria monocytogenes (strain EGD-e)  
 C/Species: listeria monocytogenes  
 C/Date: 27-Nov-2001 #sequence revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C/Accession: AF1226  
 R/Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussuberg, O.; Entian, K.D.; Fathi, H.  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A/Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria species*.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AF1226  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-183 <GLA>  
 A:Cross-references: UNIPROT:Q8YQ8; UNIPARC:UPI000055539; GB:NC\_003210; PIDN:CAC99292.1  
 A:Experimental source: strain EGD-e  
 C:Genes:  
 A:Gene: lmo1214

Alignment Scores:  
 Pred. No.: 1.95 Length: 183  
 Score: 83.00 Matches: 38  
 Percent Similarity: 37.9% Conservative: 26  
 Best Local Similarity: 22.5% Mismatches: 53  
 Query Match: 8.8% Indels: 52  
 DB: 2 Gaps: 7

US-10-768-093-4 (1-527) x AF1226 (1-183)

QY 399 CTTTAACATTTTCCCTTCATCAACATCGTAGATTAATGATGGAAG--- 343  
 DB 5 ValValIySphePheThValAaphamercGluYsGluGluAlaTyRleuAasnclumec 24  
 QY 342 -----CATCAGCGGTAATTTACTAGTCGCTTAACCCATTAATCTTCAT 301  
 DB 25 AlaGlnaNGlyTyRPhesPheGlnIySyr-----LysSerPheIySyrHis 40  
 QY 300 CATAAACGGGTAGACCGTTACT-----GGATACATTTTCAGAAACAGTTAACT 250  
 DB 41 PheGluNGlyGluProAlaIySerTyRAlaIleAspPheIySgluAasnclugly 60  
 QY 249 TCACACTAAATATTAAGCTCTT----- 226  
 DB 61 AspGluGluAlaTyRLeuThreIleuThreGluAAspAlaGlyTyRGlutThrValPheSerTyR 80  
 QY 225 -----CCTTTCACACATTTTCAACA-----AAG 202  
 DB 81 ProValIeuAenGlyAenTyRPhaArgIyAlaValAlaProGlyGluThrGlu 100  
 QY 201 AAGGCTGATTC-----GTGCTCAGCTCAAAAATTTTGAATAG 157  
 DB 101 GluAlaIlePheThAspGluThrSerLeuIleGlnLeuTyRAlaAsnIleArgIySarg 120  
 QY 156 TCGTGATACCTGGAAAGTTTATGCGCTATTCTGTCGATGCGCTGCCAATGAAG 97  
 DB 121 TrpThrIlePheGlyAlaIleValSerLeuPheIeuPhe----- 133  
 QY 96 CAAGAATTAGATTAAACCAATGTTTCTCATATAACACATTAATATTCCTAATATC 37  
 DB 134 -----ValGluLeuIleValPheGlnIleGluAAsnIleGlyIleAlaThrPhe 151  
 QY 36 ATTTTATTTGTTACATACATTCGTA 10  
 DB 152 IlePheIleIlePhePheIleVal 160

RESULT 11  
 T16580  
 Hypothetical protein K07E12.1 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C:Accession: T16580  
 R:Fulton, L.  
 submitted to the EMBL Data Library, May 1994  
 A:Description: The sequence of C. elegans cosmid K07E12.  
 A:Reference number: Z18540  
 A:Accession: T16580  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-13055 <FULL>  
 A:Cross-references: UNIPROT:Q09165; UNIPARC:UPI000017CF3A; EMBL:U00054; NID:g485140; PID

A:Experimental source: strain Bristol N2  
 C:Genes:  
 A:Gene: CESP:K07E12.1  
 A:Introns: 46/3; 85/2; 201/1; 278/2; 470/2; 817/1; 927/1; 960/3; 1265/1; 1322/1; 1478/3; 6014/3; 6159/3; 6665/2; 7266/3; 7895/3; 8669/3; 8726/3; 9803/3; 10937/3; 12234/2; 1225

Alignment Scores:  
 Pred. No.: 4.26 Length: 13055  
 Score: 83.00 Matches: 34  
 Percent Similarity: 42.2% Conservative: 23  
 Best Local Similarity: 25.2% Mismatches: 50  
 Query Match: 8.8% Indels: 28  
 DB: 2 Gaps: 6

US-10-768-093-4 (1-527) x T16580 (1-13055)

QY 144 CCAATATCAAGACTAATTTCAAAAAGTTTTCACCTGACCAAGATA-----CAG 197  
 DB 10401 ProIeuGlyThrIAspSerSerGlySerTyRlleThrGluAAspGlyGlnIeuValGlyIyS 10420  
 QY 198 CCTCTTTGGTGAATGTTGGAAGAAGAGACTTATTTATTTAGTGAACCTAACT 257  
 DB 10421 AspGluGluGlyIySProValGlyProAspGlyGlnValIeu----- 10434  
 QY 258 GTTCTGAATATGATATCCAGGTAAACGCTTACCTGTTATGATGAAGATTATGCGTTA 317  
 DB 10435 ---ProThrAspSerAlaGlyHisTyRValTyRProIle-----ThrGlyAla 10449  
 QY 318 GCAGCACTAGTAATTAATCCGCTGATGCTTCCCAATCAATATCTACAGATTGATGAG 377  
 DB 10450 AspArgGlnIleuThrThrAspAlaIleGlyIySProIleTyRSerValPheAasncl 10469  
 QY 378 AAAGGAAAAAATGTTAAAGAT-----CATGCGCAGAGCTT 416  
 DB 10470 AspGlyIleGlnIeuProThrIAspSerSerGlyTyRAlaIleGlyHisAspGlyGlnIeu 10489  
 QY 417 ACACCTATATCAAAATTAATTTTAAAGCGCTGATTAATCTACAGCGGGAATAATA 476  
 DB 10490 ValProThrGluSerThrAsnGlyValProIeuAasn-----LysAspGly 10504  
 QY 477 TCTCTGGATATATTAACGATCAGATTATGCTGTTATGTA 521  
 DB 10505 ThrProIeuProThrIAspSer-----GlyHisPheVal 10516

RESULT 12  
 G64463  
 Hypothetical protein MJ1312 - *Methanococcus jannaschii*  
 C:Species: *Methanococcus jannaschii*  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
 C:Accession: G64463  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirsch, E.F.; Weinstock, K.G.; Merrick, J.W.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurest, W.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: G64463  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-321 <BLU>  
 A:Cross-references: UNIPROT:Q58708; UNIPARC:UPI000013ATC6; GB:U67571; GB:L77117; NID:GLI  
 C:Genes:  
 A:Map position: FOR1257323-1258288

Alignment Scores:  
 Pred. No.: 3.51 Length: 321  
 Score: 81.00 Matches: 46  
 Percent Similarity: 44.6% Conservative: 29  
 Best Local Similarity: 27.4% Mismatches: 56  
 Query Match: 8.6% Indels: 37  
 DB: 2 Gaps: 10



US-10-768-093-4 (1-527) x G64463 (1-321)

OY 14 AATGTTATGATACAAATAAATGATTAAGCAATATTATGCTGATATGAGAAAC 73  
 DB 177 Serilelystrylthylglutserilellelythralagluhelelyglu----- 194  
 OY 74 AATGCTTAATTTCAATCTTCTTCAATTCGGACCCATGCGACAGAAATAGCCAC 133  
 DB 195 -----Leuanprobenlyecyetyreuleanthrproleargproproserglulys 212  
 OY 134 ---TAAAACTTCCAGATACAGACTATTTCAAAAGTTTTCGACCTGAAACACG 190  
 DB 213 Tyrilelyproprolylelelelu-Valillethrllys----- 224  
 OY 191 AATACAGCTCTTTGGTGAATAATGTTGGAAG----- 224  
 DB 225 -lleleuallalepheanenglillelelelylsanlylelyleuleuelylph 244  
 OY 225 -GAAGCA---GCTTATATTATTAGTGAATTAATCTGCTGAAAATGATCCAGGT 280  
 DB 244 egluglyasngluhelelyphelelysergluasn-----Valigluhelelyleuallal 262  
 OY 281 AACG---GTCTACCTCTGTTTATGATGAAGATTATGCGTAGACGACTATGTAATACGC 337  
 DB 262 ethserValhlsprometargglu-----Valilleysgluleuleuanslyse 280  
 OY 338 TGATGCTCCCAATCAATATCTTACCAATGTTGATGAGAAAGGAAATAATGTTAA 397  
 DB 280 rAsnlelelehelelylelleleuanslymetvalansglu---Glylylsleullelyle 299  
 OY 398 AGATCATGCTGCAGAGTTACACCTTAATCAATACTTTAAAGCGCTGAATTATAC 457  
 DB 299 uGluTyrAspelylylsVal-----PheTyrmetlylsanllely 312  
 OY 458 TAGCGGGGAAAAAATAATATCT 479  
 DB 312 sSerArgAspAsnValiser 319  
 RESULT 13  
 AF2097  
 hypothetical protein a11333 (imported) - Nostoc sp. (strain PCC 7120)  
 C/Species: Nostoc sp. PCC 7120  
 A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
 C/Accession: AF2097  
 R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saemoco, S.; Watanabe, A.; Irituguhi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A/Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Anabaena  
 A/Reference number: AB1807; MUID:21595285; PMID:11759840  
 A/Accession: AF2097  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-595 <KUN>  
 A/Cross-references: UNIPROT:Q8YUJ1; UNIPARC:UP100000CE3D3; GB:BA000019; PIDN:BA074032.1;  
 A/Experimental source: strain PCC 7120  
 C/Genetics:  
 A/Gene: a112333

# Alignment Scores:

Pred. No.: 3.93 Length: 595  
 Score: 81.00 Matches: 41  
 Percent Similarity: 38.54 Conservative: 28  
 Best Local Similarity: 22.94 Mismatches: 54  
 Query Match: 8.64 Indels: 56  
 DB: 2 Gaps: 11

US-10-768-093-4 (1-527) x AF2097 (1-595)

OY 132 ACTAAAACTCCCGATGATCAACG-----ACTATTCAAAAAGTTTTTGA 179  
 DB 184 ThrAlaArgTyrProValGlnSerProThrTyrPValValValMetGlnAspPheGlyclu 203

OY 180 CCT---GAACCAAGATACAGCCTTCT-----TTT 206

DB 204 ProValAlaProlylelleProProAlaProAsnTyrAspAlaAlaLeuPheAlaGlyAla 223

OY 207 GGTAAATGTTGGAAGAAAGAGCTTTATTTAGTGTGAACCTTAACCTCCGAA 266

DB 224 TrpIleuanslylylAlaGluGlnPheleuanslylyl-----ArgleuProGly 240

OY 267 AATGATCCAGATACAGCTGCTGTTATGATGAAGATTAT-----GGGTGAAGA 320

DB 241 AsnleuPheMetIleAsn---TrpProIlecyGelyAsnAspTyrGlyGlyValGly 259

OY 321 CGAGTAAATACCGCTGATGCT-----TCC 347

DB 260 ArgleuIleGluSerAlaAlaAlaLeuGlyAspPhePheGluGluSerArgTyrPheIleSer 279

OY 348 CAATCAATATATCTACGATGTTGATGAGAAAGGAAATAATGTTAAAGATCATGCT 407

DB 280 GlnAsnPheAlaHisPheIleGlnThrGlnleuGlyArg-----TyrGly 295

OY 408 ---GCAGAGTTTACACCTTAATCAACAAATACTTTAAAGCGCTGAATTAT----- 455

DB 296 LeuAlaGluGlnValPheProHisAlaSerSerAlaPheAlaLeuHisProTyrTyrArg 315

OY 456 -----ACTAGCGGGGAAAAAATAATCTCTCT----- 482

DB 316 GluSerArgArgLeuValGlyLeuThrThrValArgGluGlnAspIleleuProValPro 335

OY 483 -----GGAATATATACAGTACGCTTATGCTGTTCTATGTTAAAC 524

DB 336 GlyGlyGlnValAlaSerLeuPheProAspAlaIleAlaIleGlyAsnTyrAlaAsn 354

RESULT 14

C65088

hypothetical protein b3013 - Escherichia coli (strain K-12)

C/Species: Escherichia coli

C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004

C/Accession: C65088

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:9742617; PMID:9278503

A/Accession: C65088

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-309 <BLAT>

A/Cross-references: UNIPROT:Q46858; UNIPARC:UP1000013BFOE; GB:AE000383; GB:U00096; NID:9

A/Experimental source: strain K-12, substrain MG1655

C/Superfamily: Escherichia coli hypothetical protein b3013

# Alignment Scores:

Pred. No.: 3.93 Length: 309  
 Score: 80.50 Matches: 41  
 Percent Similarity: 43.24 Conservative: 32  
 Best Local Similarity: 24.34 Mismatches: 57  
 Query Match: 8.64 Indels: 39  
 DB: 2 Gaps: 10

US-10-768-093-4 (1-527) x C65088 (1-309)

OY 63 ATGAAGAAAACAATGGTTTATTTCTTCTTATTCGACGACCATGCCAGAAC 122

DB 1 MetlysllelleleuLeuPheleuAlaLeuAlaSerPheThrValHisAlaGln--- 19

OY 123 GAATAGCGACTAAAACTTCCGATGATCAACGCTATTTCAAAAGTTT----- 173

DB 20 -----ProProSerGlnThrValGluGlnThrValArgHisIle 32

OY 174 -----TTTGACCTGAACCAAGATACAGCCTCTTTTGGTGAATAATGTGAAG--- 224

DB 33 TyrGlnAsnTyrlylsSerAspAlaThrAlaProTyrPheGlyGluThrGlyGluArgAla 52



Search completed: April 10, 2006, 10:07:31  
Job time : 46 secs

```
QY 225 -----GAGAGCTTATTATTAGTGTGAAGCTTAAGTCTGTA 266
DB 53 lIeThrSerAlaArgIleGlnGlnAlaLeuThrLeuAsnAspAsnLeuThrLeuProGly 72
QY 267 AATGTATCCAGTAAAGCTGTAACCTGTTATGAT---GAAGATTATGCG----- 314
DB 73 AsnIleGlyThrLeuAspTyrAspProValCysAspCysGlnAspGlyAspLeuVal 92
QY 315 -----TTAGACGACTAGTAATACCGCTGATGCTTCCCAATCAATAATC---TAC 362
DB 93 LeuGluSerValAlaIleThrGlnThrAspAlaAspHisAlaAspAlaValAlaArgPhe 112
QY 363 CAGATTGTT---GATGAGAAAGGAAA-----AAAATGTTAAAGATCAT 404
DB 113 ArgIlePheLeuAspAspLeuGlySerThrGlnThrLeuLysMetValAlaGluSer 132
QY 405 GGTGCA-----GAGGTTACACTTAATCAACAATAACTTTAAAGCGTGAAT 452
DB 133 GlyArgTyrValIleAspAspIleValSerAsnHisGlySerValLeuGlnAlaValAsn 152
QY 453 TATACTACCGCGGAAAAAATAATATCT 479
DB 153 ---SerGluAsnGluLeuThrLeuAla 160
```

## RESULT 15

```
T47324
hypothetical protein T12K4.110 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #ext_change 09-Jul-2004
C:Accession: T47324
R:Montfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Rudd, S.; Lemcke, K.; May
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24460
A:Accession: T47324
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-951 <MON>
A:Cross-References: UNIPROT:Q9M298; UNIPARC:UPI000009FF20; EMBL:AL138640
A:Experimental source: cultivar Columbia; BAC clone T12K4
C:Genetics:
A:Map position: 3
A:Introns: 93/1: 468/3; 527/3; 565/3; 600/3; 645/3; 664/2; 696/3; 735/3; 778/1; 821/3
A:Note: T12K4.110
```

## Alignment Scores:

Pred. No.:	5.44	Length:	951
Score:	80.00	Matches:	25
Percent Similarity:	38.4%	Conservative:	8
Best Local Similarity:	29.1%	Mismatches:	29
Query Match:	8.5%	Indels:	24
DB:	2	Gaps:	3

US-10-768-093-4 (1-527) x T47324 (1-951)

```
QY 295 ACAGGCTAGACCGCTTACCTGAGTACATTTTACAGACAGTTAAGTTCACACTAAATAT 236
DB 167 ThrGlyThrValLeuCytrpGluLeuGlnAsnGlyValValSerPheThrLeuLysGly 186
QY 235 AAAGCTCCT-----TCCTTCCAAACATTTTACCAAAA 203
DB 187 ValAlaProAspThrGlyPheAsnThrSerIleValAsnIleProArgTyrSerProAsp 206
QY 202 GAAGCTGTATTCTGTGTTCAAGT----- 179
DB 207 GlyArgThrLeuAlaValProGlyLeuArgAsnAspValValMetCysAspArgPheThr 226
QY 178 GCAAAAAAATTTTGAATAGTCGTTGATACTCGGAAGTTTTCGCTATTTCTGTT 119
DB 227 GlyGluLeuPheLeu-----ArgGlyAspHisLeuGlnAlaIleCysTyr 243
QY 118 CTGGCATGGCTGCCGAAT 101
DB 244 LeuThrTyrPalaProAsn 249
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 10, 2006, 09:55:45 ; Search time 35.5 seconds  
(without alignments)  
2094.723 Million cell updates/sec

Title: US-10-768-093-4  
Sequence: 1 TTGCACATTCAGATGTTA.....TTGTTACTATGTAACCTAA 527

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p model -DEV=xlp  
-O=/abse/ABSE/EBPOC/US10768093/runat\_10042006.090303.4519/app\_query.fasta\_1  
-DB=uniprot -QRT=fastran -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPC=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.coi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abse02p  
-USER=us10768093 @CGN 1.1.466 @runat 10042006.090303.4519 -NCPU=6 -ICPU=3  
-NO MMAR -NEG\_SCORES=0 -NAIT -DSRBLCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Uniprot 05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785	88.6	154	1 F6A2_ECOLI	P53509 escherichia
2	741	78.9	154	1 F6A1_ECOLI	P53508 escherichia
3	126	13.4	167	1 F6B1_ECOLI	P53510 escherichia
4	118	12.6	167	1 F6B2_ECOLI	P53511 escherichia
5	92	9.8	440	2 O54US8_DICDI	O54US8 dictyostell
6	91	9.7	374	2 O5ZXSO_LEGPH	O5ZXSO legionella
7	90.5	9.6	318	2 O4K1L8_STRPN	O4K1L8 streptococc
8	88.5	9.5	318	2 O4K1O5_STRPN	O4K1O5 streptococc
9	88.5	9.4	1007	2 O37J18_HABPV	O37J18 heliothis a
10	88	9.4	2542	2 O54ED6_DICDI	O54ED6 dictyostell
11	87	9.3	918	2 O72P30_LEPIC	O72P30 leptospira
12	87	9.3	918	2 O8F7D7_LEBIN	O8F7D7 leptospira
13	86	9.2	2552	2 O54ED7_DICDI	O54ED7 dictyostell
14	85.5	9.1	573	2 O8IASO_PLAF7	O8IASO plasmodium
15	85.5	9.1	718	2 O9M2C6_ARATH	O9M2C6 arabidopsis
16	85	9.1	506	1 ALDH_VIBCH	P23240 vibrio chol

17	85	9.1	506	2 O7BGD6_VIBCH	O7BGD6 vibrio chol
18	84.5	9.0	437	2 O8MVR9_CLOSI	O8MVR9 clostrichs
19	84	8.9	356	2 O5X787_LEGPA	O5X787 legionella
20	84	8.9	439	1 ENO11_SCHPO	P40370 schizosacch
21	84	8.9	456	2 O4S970_TETNG	O4S970 tetraodon n
22	84	8.9	740	1 PFL_CLOPA	O4S266 clostridium
23	83.5	8.9	205	2 O8ZST2_PYRAE	O8ZST2 pyrobaculum
24	83.5	8.9	454	2 O897U6_CLOTE	O897U6 clostridium
25	83	8.8	183	2 O72OL5_LISME	O72OL5 listeria mo
26	83	8.8	183	2 O8Y7O8_LISMO	O8Y7O8 listeria mo
27	83	8.8	374	2 O5WYN8_LEGPI	O5WYN8 legionella
28	83	8.8	749	2 O81IA3_BACCR	O81IA3 bacillus ce
29	83	8.8	12268	2 O8M008_CAEEL	O8M008 caenorhabdi
30	83	8.8	13100	2 O09165_CAEEL	O09165 caenorhabdi
31	82.5	8.8	275	2 O9BDN9_PAPAN	O9BDN9 papio anubi
32	82.5	8.8	385	2 O7RKR4_PAPAYO	O7RKR4 plasmodium
33	82.5	8.8	440	2 O8A503_BACTN	O8A503 bacteroides
34	82.5	8.8	499	2 O8G8H8_PSEBP	O8G8H8 pseudomonas
35	82.5	8.8	598	2 O6A637_PROAC	O6A637 propionibac
36	82	8.7	248	2 O6H0E5_BACAN	O6H0E5 bacillus an
37	82	8.7	248	2 O63OP3_BACIZ	O63OP3 bacillus ce
38	82	8.7	248	2 O6HAS8_BACHK	O6HAS8 bacillus th
39	82	8.7	264	2 O81JU4_BACAN	O81JU4 bacillus an
40	82	8.7	268	2 O4MG01_BACCE	O4MG01 bacillus ce
41	82	8.7	279	2 O84214_BACT	O84214 uncultured
42	82	8.7	287	2 O72X92_BACCI	O72X92 bacillus ce
43	82	8.7	420	2 O6BET9_XENLA	O6BET9 xenopus lae
44	82	8.7	524	2 O6CDU2_YARLI	O6CDU2 yarrowia li
45	82	8.7	610	2 O5WV11_LEGPI	O5WV11 legionella

## ALIGNMENTS

## RESULT 1

FEA2\_ECOLI STANDARD; PRT; 154 AA.

AC P53509; 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE C56 fibribial subunit A precursor (C56 pilin).

GN Name=C56A; OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RN [1] NUCLEOTIDE SEQUENCE.

RP STRAIN=E8775; RA Wolf M.K., de Haan L.A.M., Cassels F.C., Willshaw G.A., Gestel E.C.M.,

RA Gaastria W., Warren R., Boedeker E.C.;

RU Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Fibribriae (also called pil1), polar filaments radiating

CC from the surface of the bacterium to a length of 0.5-1.5

CC micrometers and numbering 100-300 per cell, enable bacteria to

CC colonize the epithelium of specific host organs.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use as long as its content is in no way modified and this statement is not

CC removed.

DR EMBL; U04846; AAB51361.1; -; Unassigned\_DNA.

KW Fibrilia; Signal.

FT SIGNAL 1..18 Potential.

FT CHAIN 19..154 C56 fibribial subunit A.

SQ SEQUENCE 154 AA; 16940 MW; 421E223D9FASFCB8 CRC64;

## Alignment Scores:

Pred. No.: 1,99e-66  
Score: 785.00  
Percent Similarity: 100.0%

Length: 154  
Matches: 154  
Conservative: 0







OK NCBI\_TaxID=1313;  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=53/62, 688/6, and 7904/39;  
 RC Bentley S.D., Aarensen D., Mavroidi A., Saunders D., Rabinowitch E.,  
 RA Collins M., Donaghe K., Harris D., Kallio M.S., Murphy L.,  
 RA Quill M.A., Samuel G., Skovsted I.C., Barrell B.G., Reeves P.,  
 RA Parthill J., Spratt B.G.;  
 RT "Genetic analysis of the capsular biosynthetic locus from all 90  
 serotypes of *Streptococcus pneumoniae*."  
 RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, CR931665; CA133391.1; -; Genomic\_DNA.  
 DR EMBL, CR931666; CA133414.1; -; Genomic\_DNA.  
 DR EMBL, CR931664; CA133368.1; -; Genomic\_DNA.  
 KW Transferase.  
 SQ SEQUENCE 318 AA; 37207 MW; 26ADD45FBFBED25 CRC64;  
 Alignment Scores:  
 Pred. No.: 8.54 Length: 318  
 Score: 90.50 Matches: 43  
 Percent Similarity: 42.8% Conservative: 19  
 Best Local Similarity: 29.7% Mismatches: 58  
 Query Match: 9.6% Indels: 25  
 DB: 2 Gaps: 6  
 US-10-768-093-4 (1-527) x Q4K105\_STRPN (1-318)  
 QY 60 TATATGAAGAAACAATTGTTAATCTA-----ATT 92  
 DB 17 TyrLeuLysSerIleGluSerValIleuAnGlnThrTyGlnAenIleGluIleu 36  
 QY 93 CTTCGCTTATTCGAGCCATGCCAGAAAGATAAGCACTTAACCTTCCAGTATCA 152  
 DB 37 LeuValAspGlySer-----ThraSerSerGlyLys----- 48  
 QY 153 ACGACTATTTCAAAAGTTTTCGACCTGACCAACGATACAGACTTCTTTGTGA 212  
 DB 49 -----IleCyGluSerPheSerLysValAspProArgIleArgValPheHisLysGlu 66  
 QY 213 AATGTGGAAGGAAGAGCTTATTATTAGTGGAAGCTTAACGTCTCGAAATGTA 272  
 DB 67 AenGlyGlyLeuSerAspAlaArgAenPheGlyIleGluGlnMetLysGlyGlnTyVal 86  
 QY 273 TCCAGGTACGCGTCTACCTGTTATGATGAAGATTAT-----GGGTAGAGAGACTA 326  
 DB 87 AlaPheIleAspSerAspAspTyrlleSerLysAspTyValTrpLysLeuTySerSer 106  
 QY 327 GTAATACCGCTGATGCTTCCCAATCAATATC--TACCAAGTTGTTATGAGAAAGG 383  
 DB 107 IleLysAenAenAspSerGluValSerIleCySerPheLeuLeuValAspGluLysGly 126  
 QY 384 AAAAAAATG-----TTAAAGATCATGTGCGAGAGTTACACCTAATCAACAATA 434  
 DB 127 GluLysIleLysAspGluLeuLeuAspSerGlyLysIleCySerLeuThrGlyGlnIle 146  
 QY 435 ACTTTAAAGCGCTG 449  
 DB 147 LeuGluLysValLeu 151  
 RESULT 8  
 Q4K105\_STRPN PRELIMINARY; PRT; 318 AA.  
 AC Q4K105;  
 DT 13-SEP-2005 (TREMBLrel. 31, Created)  
 DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)  
 DE Putative glycosyl transferase.  
 GN ORFNames=SPC15A\_0013;  
 OS *Streptococcus pneumoniae*.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus;  
 OK NCBI\_TaxID=1313;  
 RN [1]

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=389/39;  
 RA Bentley S.D., Aarensen D., Mavroidi A., Saunders D., Rabinowitch E.,  
 RA Collins M., Donaghe K., Harris D., Kallio M.S., Murphy L.,  
 RA Quill M.A., Samuel G., Skovsted I.C., Barrell B.G., Reeves P.,  
 RA Parthill J., Spratt B.G.;  
 RT "Genetic analysis of the capsular biosynthetic locus from all 90  
 serotypes of *Streptococcus pneumoniae*."  
 RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, CR931663; CA133345.1; -; Genomic\_DNA.  
 KW Transferase.  
 SQ SEQUENCE 318 AA; 37209 MW; 0190E7D08D6AAE33 CRC64;  
 Alignment Scores:  
 Pred. No.: 10.7 Length: 318  
 Score: 89.50 Matches: 42  
 Percent Similarity: 42.8% Conservative: 20  
 Best Local Similarity: 29.0% Mismatches: 58  
 Query Match: 9.5% Indels: 25  
 DB: 2 Gaps: 6  
 US-10-768-093-4 (1-527) x Q4K105\_STRPN (1-318)  
 QY 60 TATATGAAGAAACAATTGTTAATCTA-----ATT 92  
 DB 17 TyrLeuLysSerIleGluSerValIleuAnGlnThrTyGlnAenIleGluIleu 36  
 QY 93 CTTCGCTTATTCGAGCCATGCCAGAAAGATAAGCACTTAACCTTCCAGTATCA 152  
 DB 37 LeuValAspGlySer-----ThraSerSerGlyLys----- 48  
 QY 153 ACGACTATTTCAAAAGTTTTCGACCTGACCAACGATACAGACTTCTTTGTGA 212  
 DB 49 -----IleCyGluSerPheSerLysValAspProArgIleArgValPheHisLysGlu 66  
 QY 213 AATGTGGAAGGAAGAGCTTATTATTAGTGGAAGCTTAACGTCTCGAAATGTA 272  
 DB 67 AenGlyGlyLeuSerAspAlaArgAenPheGlyIleGluGlnMetLysGlyGlnTyVal 86  
 QY 273 TCCAGGTACGCGTCTACCTGTTATGATGAAGATTAT-----GGGTAGAGAGACTA 326  
 DB 87 AlaPheIleAspSerAspAspTyrlleSerLysAspTyValTrpLysLeuTySerSer 106  
 QY 327 GTAATACCGCTGATGCTTCCCAATCAATATC--TACCAAGTTGTTATGAGAAAGG 383  
 DB 107 IleLysAenAenAspSerGluValSerIleCySerPheLeuLeuValAspGluLysGly 126  
 QY 384 AAAAAAATG-----TTAAAGATCATGTGCGAGAGTTACACCTAATCAACAATA 434  
 DB 127 GluLysIleLysAspGluLeuLeuAspSerGlyLysIleCySerLeuThrGlyGlnIle 146  
 QY 435 ACTTTAAAGCGCTG 449  
 DB 147 LeuGluLysValLeu 151  
 RESULT 9  
 Q37318\_HABPV PRELIMINARY; PRT; 1007 AA.  
 AC Q37318;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Spheroidin.  
 OS *Heliothis armigera entomopoxvirus* (HABPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae.  
 OK NCBI\_TaxID=10290;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=98062131; PubMed=9400959;  
 RA Srikantha A., Osborne R.J., Dall D.J.;  
 RT "Mapping of the *Heliothis armigera entomopoxvirus* (HABPV) genome, and  
 analysis of genes encoding the HABPV spheroidin and nucleoside  
 triphosphate phosphohydrolase I proteins."

RL J. Gen. Virol. 78:3115-3123(1997).  
 RN (2)  
 RP NUCLEOTIDE SEQUENCE.  
 RA Srikenchana A., Osborne R.J., Dail D.J.  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF019224; AAB96623.1; -; Genomic\_DNA.  
 DR InterPro; IPR008843; Spheroidin.  
 DR Pfam; PF05541; Spheroidin; 1.  
 SQ SEQUENCE 1007 AA; 11538 MW; 3352847F565B6D8 CRC64;  
  
 Alignment Scores:  
 Pred. No.: 14.5 Length: 1007  
 Score: 88.50 Matches: 33  
 Percent Similarity: 42.2% Conservative: 29  
 Best Local Similarity: 22.4% Mismatches: 46  
 Query Match: 9.4% Indels: 39  
 DB: 2 Gaps: 6  
  
 US-10-768-093-4 (1-527) x O37318\_HABPV (1-1007)  
 QY 54 TGGTGTATATGAAAGAAATGTTAATTCTATCTGCTCATTCGGCAGCCAT 113  
 DB 99 TTPLYSTYVALSERK-----LeuLeuLeuLalaenVal--SerHis 112  
 QY 114 GCGAAGACAGATATACGACTAAATCTCCAGATCAACGACTATTTCAAAAAGTTT 173  
 DB 113 AapAenValIyetyrIyLeuLalaenTytyrLeu-----ThreLeuApenLyshis 130  
 QY 174 TTTCACCTGAAACGAAATACAGCT-----TCITTTGGTAAATGTGGAAG 224  
 DB 131 IleybProLyshIleIleAapGlnProLeuPheIleTytyrValAaSpAspleuGlnHis 150  
 QY 225 GAAGAGCTTTA-----TTATTAGTGACACTTAATCTGCTGGAATGTA 272  
 DB 151 TYRGLYLeuIleThyIyLeuLalaenIleTytyrAaAaSpAenValaenGlnAaSpala 170  
 QY 273 TCCAGCTACGCTTACCTGCTTTAT----- 299  
 DB 171 SerPheIleThyValPheProGlnTytyrAlaTytyrIleHisMetGlytyrGlyValTytyrIle 190  
 QY 300 -----GATGACATTAATGGTTAGCA 320  
 DB 191 AaenGlytyrSerThyPheAaPValThyThyAaSpalaThyAaAaIleAaenIleAaenPheAa 210  
 QY 321 CGACTAGTAATATACCGCTATGCTTCCCAATCAATATATCATCAGATGTTGATGAGAA 380  
 DB 211 LySerValaAenIleIleValaSerPheLeuAaPleTytyrGlyValaAaAaAaenGln 230  
 QY 381 GGGAAAAAATGTTAAAGAT 401  
 DB 231 GILYGLYLeuLeuLeuLyAaP 237  
  
 RESULT 10  
 Q54ED6\_DICDI PRELIMINARY; PRT; 2542 AA.  
 AC 054ED6;  
 DT 13-SEP-2005 (TREMBlrel. 31, Created)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=DOB0219680;  
 GN Dicyosetellium discoidium (slime mold).  
 OC Eukaryota; Mycetozoa; Dicyosetellida; Dicyosetellium.  
 NC NCBI\_TaxId=44689;  
 RN (1)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=XX4;  
 RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,  
 RA Sugeng R., Berthman M., Song J., Olsen R., Szafianski K., Xu Q.,  
 RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,  
 RA Benlier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
 RA Plicher K., Chen G., Saunders D., Sodergren E., Davis P.,  
 RA Kernionou A., Nie X., Hall N., Anford C., Hemphill L., Baeon N.,

RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,  
 RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
 RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
 RA Hauser H., James K., Oules M., Mohan M.B., Saito T., Buchrieser C.,  
 RA Martorep A., Felder M., Phangavelu M., Johnson D., Knights A.,  
 RA Louisged H., Mungall K., Oliver K., Price C., Qual M.A., Sanders M.,  
 RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D.,  
 RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
 RA Sugano S., White B., Walker D., Woodward J., Winkler T., Tanaka Y.,  
 RA Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,  
 RA Chaisoim R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,  
 RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.,  
 RT Nature 0:0-0(2005).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAF10100257; BAL61826.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 2542 AA; 28853 MW; D28121989617A16 CRC64;  
  
 Alignment Scores:  
 Pred. No.: 17.4 Length: 2542  
 Score: 88.00 Matches: 38  
 Percent Similarity: 39.2% Conservative: 27  
 Best Local Similarity: 22.9% Mismatches: 59  
 Query Match: 9.4% Indels: 42  
 DB: 2 Gaps: 6  
  
 US-10-768-093-4 (1-527) x Q54ED6\_DICDI (1-2542)  
 QY 72 ACAATGGTTAATATCTAATCTGCTCATTCGGCAGCCATGCGAAGAAATAGCG 131  
 DB 2357 SerIleGlyLeuProThyIleCythrAaenTytyrGlyLeuIleGlySerThyGlyPheVal 2376  
 QY 132 ACTAAATCTCCAGTTCACAGACTATTTCAAAAAGTTTTCACCT----- 182  
 DB 2377 SerArgAaenGlySerValaIleAaLeuLeuSerGlyGlyPheIleProIleSerAla 2396  
 QY 183 -----GAACCAAGAAATACAGCCTCTTTTGTGGAATGTTGA 221  
 DB 2397 AaenLeuIleLeuGlyThyLeuAaSpLeuGlnLeuGln-----AaenGlnAaGln 2412  
 QY 222 AAGGAAGAGCTTATTAATTAATGATGTAATTAATCTGCTGAAAATGATATCCAGTA 281  
 DB 2413 SerSerAaenLeuIleLeuSerAaAaPheThySerLeuAaenGlyLeuProGlnLyb 2432  
 QY 282 ACGCTTACCTGCTTATGATGAAATATATGCTTATGACACATGTAATACCGCTGAT 341.  
 DB 2433 SerLeuIleSerTytyrPheAaP----- 2439  
 QY 342 GCTCCCAATCAATATATCTACAGAT-----GTGATGAGAAAGGAAAAA 389  
 DB 2440 -----TytyrIleAaenIleAaenGlnGlyAaenGlySerTytyrSer 2453  
 QY 390 ATGTTAAAGATCATGTCAGAGGTTACACTTAATCAACAATATCTTTAAAGCGCTG 449  
 DB 2454 LeuLeuLyAaPAsPAsPAsPValaGlyLeuThyValaAaPValaAaenGlyThyPheTytyrSer 2473  
 QY 450 AATTATCTACGCGGGGAA-----AAAAAATATCTCTGGAATATTAAGATCAGTTATG 506  
 DB 2474 GILYLeuSerThyAaPleIleuLyLeuAaen-----LyAaPleIleu 2489  
 QY 507 GTTGTTACTATGTAAC 524  
 DB 2490 VALAaPTyTYGLYIleAaP 2495  
  
 RESULT 11  
 Q72P30 LEPIIC  
 ID 072P30 LEPIIC PRELIMINARY; PRT; 918 AA.  
 AC 072P30;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)



DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Membrane carboxypeptidase.  
 GN OrderedLocustNames=LI12546;  
 OS Leptospira interrogans (serogroup Icterohaemorrhagiae / serovar  
 Copenhagen)  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxId=44275;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Fiocruz LI-110;  
 RX PubMed=15028702; DOI=10.1128/JB.186.7.2164-2172.2004;  
 RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,  
 Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,  
 Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carver H.,  
 Coutinho L.L., Degraive M.M., Dellagostin O.A., El-Dorry H.,  
 Ferro E.S., Ferro M.T.T., Furlan L.R., Gambellini M., Gigliotti E.A.,  
 Goeb-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,  
 Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,  
 Kuramae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,  
 de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A.,  
 Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,  
 Camargo L.E.A., Kitajima J.P., Setubal J.C., Van Sluys M.A.,  
 "Comparative genomes of two Leptospira interrogans serovars reveals  
 novel insights into physiology and pathogenesis.";  
 RL J. Bacteriol. 186:2164-2172(2004).  
 DR EMBL, AB017297; AAS71206.1; -; Genomic DNA.  
 DR GO:GO:0009274; C:cell wall (sensu Bacteria); IEA.  
 DR GO:GO:0004180; F:carboxypeptidase activity; IEA.  
 DR GO:GO:0003824; F:catalytic activity; IEA.  
 DR GO:GO:0008658; F:penicillin binding; IEA.  
 DR GO:GO:0009252; P:peptidoglycan biosynthesis; IEA.  
 DR InterPro: IPR001264; Glyco\_trans\_51.  
 DR InterPro: IPR001460; Pencl\_bind\_1pept.  
 DR Pfam: PF00912; Transgly\_1.  
 DR Pfam: PF00905; Transpeptidase; 1.  
 DR ProDom: PD001895; Glyco\_trans\_51; 1.  
 DR Carboxypeptidase; Complete proteome.  
 SO SEQUENCE 918 AA; 102807 MW; 2D9E512BC396960 CRC64;  
 Alignment Scores:  
 Pred. No.: 20.1 Length: 918  
 Score: 87.00 Matches: 37  
 Percent Similarity: 43.4% Conservative: 25  
 Best Local Similarity: 25.9% Mismatches: 57  
 Query Match: 9.3% Indels: 24  
 DB: 2 Gaps: 6  
 US-10-768-093-4 (1-527) x Q72P30\_LEPIC (1-918)  
 QY 132 ACTAAAACTCCGATCAAGCACTATTTCAAAAAGTTTTCACCTGAACCAAGA 191  
 DB 478 ThrGlyGlyPheSerIleYrThrThrValSerGluProValGlnAlaGluLeuProlys 497  
 QY 192 ATACAGCCTTCTTGGTGAAGAAATGTTGAAAGAGAGAGCTTTATTTAGTGTGAAC 251  
 DB 498 IleValLysAsnYrValAspAsnValGlnLysAsnGly---LeuValArgLysThrArg 516  
 QY 252 TTAAGTCTCCGAAAGATATATCCAGGTAAACGGTACCTGTTATATGATGAATAT 311  
 DB 517 LeuThrAspAsnLysAsnSerSerGluThrAlaValPheArgArgTyrIleGlnAspLeu 536  
 QY 312 -----GGGTAGAGCACTAGTAATATAC-----GCTGATGCTTCC 347  
 DB 537 SerProAlaLeuGluLeuPheIleAspThrAspSerPheGlyGlyGlnAsnGluSerGly 556  
 QY 348 CAATCAATATATACAGATGTGTTGATGAGAAAGGAGAAAGATGTTAAAGATCATGCT 407  
 DB 557 LeuGlnValAlaLeuValAlaValAspProSerThrGlyGlnIleLeuLeuMetHisGly 576  
 QY 408 GGAGAGGTTACACCTAATACAAAGAAATATTTTAAAGCG-----CTGAATTATATCT 458  
 DB 577 GlySer-----GluPheLysAlaAspAsnGlnLeuAspArgThr 589

QY 459 AGCGGGGAAAAA-----ATATCTCGAATATATACGATCAGGTT 503  
 DB 590 ThrAlaMetLysArgGlnThrGlySerSerIleLysProIleLeuYrSerAlaAlaIle 609  
 QY 504 ATGTTGCT 512  
 DB 610 GluThrGly 612  
 RESULT 12  
 Q8F7D7\_LEPIN  
 ID Q8F7D7\_LEPIN PRELIMINARY; PRT; 918 AA.  
 AC Q8F7D7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Penicillin-binding protein 1A (EC 3.4.4.-) (EC 2.4.2.-).  
 GN OrderedLocustNames=LA1009;  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxId=173;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;  
 RX MEDLINE=22598143; PubMed=12712204; DOI=10.1038/nature01577;  
 RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,  
 Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,  
 Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,  
 Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,  
 Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,  
 Saint Glions I., Somerville R.L., Wen Y.-W., Shi M.-H., Chen Z.,  
 Xu J.-G., Zhao G.-P.;  
 RA "Unique physiological and pathogenic features of Leptospira  
 RT Interrogans revealed by whole-genome sequencing.";  
 RL Nature 422:888-893(2003).  
 DR EMBL, AB011285; AAM48208.1; -; Genomic DNA.  
 DR GO:GO:0009274; C:cell wall (sensu Bacteria); IEA.  
 DR GO:GO:0003824; F:catalytic activity; IEA.  
 DR GO:GO:0008658; F:penicillin binding; IEA.  
 DR GO:GO:0009252; P:peptidoglycan biosynthesis; IEA.  
 DR InterPro: IPR001264; Glyco\_trans\_51.  
 DR InterPro: IPR001460; Pencl\_bind\_1pept.  
 DR Pfam: PF00912; Transgly\_1.  
 DR Pfam: PF00905; Transpeptidase; 1.  
 DR ProDom: PD001895; Glyco\_trans\_51; 1.  
 KW Complete proteome.  
 SO SEQUENCE 918 AA; 102818 MW; 5EBA6FD48DC58326 CRC64;  
 Alignment Scores:  
 Pred. No.: 20.1 Length: 918  
 Score: 87.00 Matches: 37  
 Percent Similarity: 43.4% Conservative: 25  
 Best Local Similarity: 25.9% Mismatches: 57  
 Query Match: 9.3% Indels: 24  
 DB: 2 Gaps: 6  
 US-10-768-093-4 (1-527) x Q8F7D7\_LEPIN (1-918)  
 QY 132 ACTAAAACTCCGATCAAGCACTATTTCAAAAAGTTTTCACCTGAACCAAGA 191  
 DB 478 ThrGlyGlyPheSerIleYrThrThrValSerGluProValGlnAlaGluLeuProlys 497  
 QY 192 ATACAGCCTTCTTGGTGAAGAAATGTTGAAAGAGAGAGCTTTATTTAGTGTGAAC 251  
 DB 498 IleValLysAsnYrValAspAsnValGlnLysAsnGly---LeuValArgLysThrArg 516  
 QY 252 TTAAGTCTCCGAAAGATATATCCAGGTAAACGGTACCTGTTATATGATGAATAT 311  
 DB 517 LeuThrAspAsnLysAsnSerSerGluThrAlaValPheArgArgTyrIleGlnAspLeu 536  
 QY 312 -----GGGTAGAGCACTAGTAATATAC-----GCTGATGCTTCC 347  
 DB 537 SerProAlaLeuGluLeuPheIleAspThrAspSerPheGlyGlyGlnAsnGluSerGly 556



```

QY 348 CATCATATATCTACCAAGTTGTCAGAAAGGAAAAATGTTAAAGATCATGCT 407
DB 557 LeuGlnValAlaLeuValAlaValAlaProSerThrGlyGluLeuLeuMetHisGly 576
QY 408 GCAGAGCTTACACCTTAATCAACAAATACCTTTAAAGCG-----CTGATTTACT 458
DB 577 GlySer-----GluPheValAlaPheAspGlnLeuAspArgThr 589
QY 459 AGCGGGAAGAAAAA-----ATATCTCTGGAATATATATACATCAGTT 503
DB 590 ThrAlaMetLeuArgGlnThrGlySerSerIleLeuProIleLeuYrSerAlaIle 609
QY 504 ATGTTGCT 512
DB 610 GluThrGly 612

RESULT 13
OS4ED7_DICDI PRELIMINARY; PRT; 2552 AA.
AC OS4ED7;
DT 13-SEP-2005 (TReMBLrel. 31, Created)
DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
DE Hypothetical protein.
CN ORFName=DD80191589;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sungang R., Berriman M., Song J., Olsen R., Szatranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pijcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kethornou A., Nie X., Hall N., Anjard C., Hemphill L., Bacon N.,
RA Fardocher P., Desany B., Juet E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Moutier T., Pain A., Harper D., Lindsay R.,
RA Hauser H., James K., Oulles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulès H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabbittowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.,
RA RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
SQ EMBL: AAF01000257; EAL61791.1; -; Genomic_DNA.
KM Hypothetical protein.
SQ SEQUENCE 2552 AA; 289091 MW; 0515D8D76EBCB4FA CRC64;

Alignment Scores:
Pred. No.: 27.1 Length: 2552
Score: 86.00 Matches: 44
Percent Similarity: 35.4% Conservative: 20
Best Local Similarity: 24.3% Mismatches: 45
Query Match: 9.2% Indels: 72
DB: 2 Gaps: 7

US-10-768-093-4 (1-527) x OS4ED7_DICDI (1-2552)
QY 72 ACAATTGGTTTAATCTAATCTTCTGCTTCCGACCATGCCAGAACAGAAATAGCG 131
DB 2366 SerIleGlyLeuProThrIleCysThrAsnTyGlyLeuIleGlnSerThrGlyPheVal 2385
QY 132 ACTAAAACTTCCAGTATCAAGCATATTTCAAAAAGTTTTCACCTGACCAACAGCA 191

```

```

DB 2366 SerAspAsnGlnSerValAlaAlaLeuSer----- 2396
QY 192 ATACAGCCTTCTTTGCGTAAATGTTGAAAGAGAGCTTTATTTACTGTGAC 251
DB 2397 -----GlyGlnGlyLeuLeuProIleSerThrAsn 2406
QY 252 TTAAGTCTCTGAAATGTATCCAGGTAAACGCTACCTGTTTATGATGAGATTA 311
DB 2407 LeuIle----- 2408
QY 312 GGGTTAGG-----CGACTAGTAATACCGGTAGTCTTCCCAATCAATATAC 359
DB 2409 --LeuGlyThrLeuAspLeuGlnLeuGlnAlaGlnSerSerAsnLeuLeu 2427
QY 359 ----- 359
DB 2428 SerAsnPheAsnPheThrSerLeuAsnGlyLeuProGlnYrSerLeuIleSerLysPhe 2447
QY 360 ---TACCAATT-----CTGATGAGAAAGGAAAAATGTTAAAGATCAT 404
DB 2448 AspTyrglnIleAsnIleAsnGlnGluAsnGlnYrSerLysSerLeuLeuAspAsp 2467
QY 405 GGTGACAGGTTTACACCTTAATCAACAAATTAATTAAAGCGCTGATTAATAC 464
DB 2468 AsnValGluLeuThrValAspGlnLeuIleThrPheLysIleSerGluLeuSerThr 2487
QY 465 GAA--AAAAAATATCTCTCGGAATATTAACGATCAGTTAGTTGTTACTATGTA 521
DB 2488 AspIleLeuLysLeuAsn-----LysAspIleIleLeuValAspTyrcIlyVal 2503
QY 522 AAC 524
DB 2504 Asp 2504

RESULT 14
OS1AS0_PLAF7 PRELIMINARY; PRT; 573 AA.
ID OS1AS0;
AC OS1AS0;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Peptinogen, putative (EC 3.4.23.1).
GN Name=PF08_0108;
OS Plasmodium falci-parum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.,
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL844507; CAD51290.1; -; Genomic_DNA.
DR HSSP: P00797; 2REN.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0004194; F:pepsin A activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR001461; Peptidase A1.
DR InterPro: IPR001969; Pept_Asp_AS.
DR Pfam: PF00026; Asp; 1.
DR PRINTS: PR00792; PEPsin.
DR PROSITE: PS00141; ASP_PROTASE; UNKNOWN_1.
KM Hydrolase.
SQ SEQUENCE 573 AA; 65114 MW; 1381CB0D4519A54F CRC64;

Alignment Scores:
Pred. No.: 27 Length: 573
Score: 85.50 Matches: 46
Percent Similarity: 41.4% Conservative: 26
Best Local Similarity: 26.4% Mismatches: 65
Query Match: 9.1% Indels: 37
DB: 2 Gaps: 8

```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

```
Run on:      April 10, 2006, 10:06:30 ; Search time 6.7 Seconds
              (without alignments)
              1300.599 Million cell updates/sec
```

```
Title: US-10-768-093-4
Perfect score: 939
Sequence: 1 TTGACACATTACGATGTTA.....TTGGTACTACTGTAACATA 527
```

Scoring table:	
BLOSUM62	
Xgapop 10.0 ,	Xgapext 0.5
Ygapop 10.0 ,	Ygapext 0.5
Fgapop 6.0 ,	Fgapext 7.0
Delop 6.0 ,	Delext 7.0

```
Searched:      572060 seqs, 82675679 residues
Total number of hits satisfying chosen parameters: 1144120
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Command line parameters: -DEV=xlp
-MODEL=frame-n2p.model
-Os/abs/ABSSMEM.spool/US10766093/runat.10042006.09307.4612/app.qcv.fasra.1
-B/seq/abs/Patents_AA_QSMT/fasraen_SUFFIX-n2p.ra1-MINMATCH=0.1-LOOPTC=0
-LOOPTC=0-INITIS=bits-START1=1-END=1-MATRX=10lsum62-PRANS=numa00.cdt
-LIST=445-DOCNAL=200-THR SCORE=pcp-THR MAX=100-THR MIN=0-ALIGN=15
-MODE=LOCAL-OUTTM=ptco-NONMEXC=HEAPSIZE=500-MINLEN=10-MAXLEN=2000000000
-HOST=abse802-user=US10766093-CAIT=1.71-@runat.10042006.09307.4612-NCPU=6
-ICPU=0-NO MMAP-NEG SCORES=0-WAIT-DBSPBLX=100-LOGNLOG-DEV TIMEOUT=120
-MARN TIMEOUT=30-THREADS=1-XGAPOP=10-XGAPEXT=0.5-FGAPOP=6-FGAEXT=7
YGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXEXT=10

```

```
Database :
Issued Parents AA:*
1 : /cgn2.6/plodacta.1/1aa/5 COMB.pep:*
2 : /cgn2.6/plodacta.1/1aa/6 COMB.pep:*
3 : /cgn2.6/plodacta.1/1aa/H- COMB.pep:*
4 : /cgn2.6/plodacta.1/1aa/PTCUS COMB.pep:*
5 : /cgn2.6/plodacta.1/1aa/RE COMB.pep:*
6 : /cgn2.6/plodacta.1/1aa/backfill.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	86	9.2	378	US-09-107-532A-4742	Sequence 4742, App
2	84	8.9	740	US-09-353-872A-25	Sequence 25, Appl
3	84	8.9	740	US-09-072-433-30	Sequence 30, Appl
4	82	8.7	330	US-09-248-796A-14853	Sequence 14853, App
5	81	8.6	768	US-09-114-000C-6467	Sequence 6467, App
6	77	8.2	345	US-09-558-092-314	Sequence 314, Appl
7	77	8.2	483	US-09-252-91A-21015	Sequence 21015, App
8	76.5	8.1	186	US-09-543-681A-6955	Sequence 6955, App
9	76.5	8.1	308	US-09-248-796A-18392	Sequence 18392, App
10	76.5	8.1	1076	US-09-114-001C-4037	Sequence 4037, Appl
11	76	8.1	218	US-09-328-352-6886	Sequence 6886, App
12	76	8.1	527	US-09-252-91A-23709	Sequence 23709, App

13	75.5	8.0	603	2	US-09-902-540-11397	Sequence 11397, Appl
14	75.5	8.0	1184	2	US-09-541-782-2	Sequence 2, Appl
15	75.5	8.0	1184	2	US-09-723-890-2	Sequence 2, Appl
16	75.5	8.0	1184	2	US-10-270-085-2	Sequence 2, Appl
17	75	8.0	130	2	US-09-134-000C-5020	Sequence 5020, Appl
18	74.5	7.9	400	2	US-09-710-219-2774	Sequence 2774, Appl
19	74.5	7.9	497	2	US-09-134-001C-5114	Sequence 5114, Appl
20	74	7.9	170	1	US-08-913-477-21	Sequence 21, Appl
21	74	7.9	302	2	US-09-328-352-7673	Sequence 7673, Appl
22	74	7.9	467	2	US-09-246-796A-17536	Sequence 17536, Appl
23	74	7.9	501	1	US-08-913-477-23	Sequence 23, Appl
24	73.5	7.8	125	2	US-09-328-352-2033	Sequence 5033, Appl
25	73.5	7.8	584	2	US-09-107-532A-4564	Sequence 4564, Appl
26	73.5	7.8	1006	2	US-09-710-279-154	Sequence 154, Appl
27	73.5	7.8	1806	2	US-09-919-497-56	Sequence 56, Appl
28	72.5	7.7	465	2	US-09-769-787-164	Sequence 164, Appl
29	72	7.7	149	2	US-09-113-000C-4976	Sequence 4976, Appl
30	72	7.7	1005	2	US-09-944-016-6079	Sequence 6079, Appl
31	72	7.7	1008	2	US-09-949-016-1359	Sequence 7359, Appl
32	71.5	7.6	84	2	US-09-270-767-7692	Sequence 7692, A
33	71.5	7.6	84	2	US-09-270-767-52909	Sequence 52909, A
34	71.5	7.6	375	2	US-09-270-767-11678	Sequence 11678, A
35	71.5	7.6	375	2	US-09-270-767-66895	Sequence 46895, A
36	71.5	7.6	399	2	US-09-583-110-2754	Sequence 2754, Ap
37	71.5	7.6	560	2	US-08-983-045-4	Sequence 4, Appl
38	71.5	7.6	1667	2	US-09-134-000C-6740	Sequence 6740, Ap
39	71	7.6	289	2	US-09-107-532A-7169	Sequence 7169, Ap
40	70.5	7.5	305	2	US-09-328-352-6069	Sequence 6069, Ap
41	70.5	7.5	323	2	US-09-538-092-113	Sequence 213, Appl
42	70.5	7.5	357	2	US-09-103-331-11	Sequence 11, Appl
43	70.5	7.5	357	2	US-09-631-594-40	Sequence 50, Appl
44	70.5	7.5	452	2	US-09-538-092-371	Sequence 371, Appl
45	70.5	7.5	452	2	US-09-107-532A-4858	Sequence 4858, Ap

## ALIGNMENTS

RESULT 1  
US-09-107-532A-4742  
; Sequence 4742, Application US/09107532A

GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310

```
? COMPUTER READABLE FORM:  
? MEDIUM TYPE: CD-ROM ISO9660  
? COMPUTER: PC  
? OPERATING SYSTEM: <Unknown>  
? SOFTWARE: ASCII
```

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,  
 FILING DATE: 30-Jun-1998 .  
 PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
NAME: Arianello Pamela  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 893-5007  
TELEFAX: (781) 893-8277

INFORMATION FOR SEQ ID NO: 4742:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 378 amino acids  
TYPE: amino acid  
MOLECULE TYPE: linear  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (8) LOCATION 1...378  
SEQUENCE DESCRIPTION: SEQ ID NO: 4742:  
US-09-107-532A-4742

Alignment Scores:  
Pred. No.: 0.0337 Length: 378  
Score: 86.00 Matches: 39  
Percent Similarity: 42.7% Conservative: 25  
Best Local Similarity: 26.0% Mismatches: 56  
Query Match: 9.2% Indels: 30  
Gaps: 8

US-10-768-093-4 (1-527) x US-09-107-532A-4742 (1-378)

QY 90 ATTCTGCTTCATTCGCGAGCCATGCGAGAGAAATAGCGACTTAAACTTCCAGTA 149  
:::|||||:::|||||:::|||||:::|||||:::|||||  
158 VALLEALALALALAGLYALAGLYTHRGINGLYASPTY-THRGlyThrsrhe----- 175  
:::|||||:::|||||:::|||||:::|||||:::|||||  
150 TCAAGCACTATTTCAGAAAGTTTTCGACCTGAAACGAGATACAGCTTCTTTGCT 209  
:::|||||:::|||||:::|||||:::|||||:::|||||  
176 ---ThrsrlelglYHlglYArGPhetrPrGluAlglYAlglInPrAlallegly 194  
:::|||||:::|||||:::|||||:::|||||:::|||||  
210 GAAATGTTGGAAG---GAGGAGCTTATATTATTAGTGAACCTGACTGCTGGA 266  
:::|||||:::|||||:::|||||:::|||||:::|||||  
195 Lye---ValglYlYThrgInglInValglInglAlalYValglInValleuPrGlu 213  
:::|||||:::|||||:::|||||:::|||||:::|||||  
267 AATGATCCAGTACGCTTAC-----GAGATTAAGCTTACGAGACTAGTAAT 296  
:::|||||:::|||||:::|||||:::|||||:::|||||  
214 ThrleGlulYglInVallelglInAlameArGsrAlahlsPrGlyrGlulGluProAla 233  
:::|||||:::|||||:::|||||:::|||||:::|||||  
297 TATGAT-----GAGATTAAGCTTACGAGACTAGTAAT 332  
|||||:::|||||:::|||||:::|||||:::|||||  
234 TyArPrleuPheAlalaleaPrGluProValGlulmeCPhelYleuGlYArGVal----- 251  
:::|||||:::|||||:::|||||:::|||||:::|||||  
333 ACCGCTGATGCTCCCAATCAATTAATCTACAGATTGTGAGAGAAAGGAAATG 392  
:::|||||:::|||||:::|||||:::|||||:::|||||  
252 ---GlYglulPrGluInglulPrGluInglulAlaPheValGlulInValleYglulAla 270  
:::|||||:::|||||:::|||||:::|||||:::|||||  
393 TTAAAGATCATGTGCGAGAGGTTACA-----CCTAATCAACAAATTAATTAA-- 443  
:::|||||:::|||||:::|||||:::|||||:::|||||  
271 PheInglulAlaPrGluYleuArGlleValglInPrGluYAsnAlalYserseVallybArG 290  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 444 ---GGCTGAATTATCTACCGGGGAAAAA 470  
|||||:::|||||:::|||||:::|||||:::|||||  
Db 291 lleaAlaleCyeGlYlYserGlYglulYs 300  
|||||:::|||||:::|||||:::|||||:::|||||

RESULT 2  
US-09-323-872A-25  
Sequence 25, Application US/09323872A  
Patent No. 6395539  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter  
TITLE OF INVENTION: Compositions and Methods for Bioremediation  
FILE REFERENCE: OHU-03640  
CURRENT APPLICATION NUMBER: US/09323, 872A  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: 09/072, 433  
NUMBER OF SEQ ID NOS: 58  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 25  
LENGTH: 740

TYPE: PRT  
ORGANISM: Clostridium pasteurianum  
US-09-323-872A-25

Alignment Scores:  
Pred. No.: 0.0849 Length: 740  
Score: 84.00 Matches: 39  
Percent Similarity: 38.6% Conservative: 25  
Best Local Similarity: 23.5% Mismatches: 68  
Query Match: 8.9% Indels: 34  
Gaps: 7

US-10-768-093-4 (1-527) x US-09-323-872A-25 (1-740)

QY 93 CTGCTTCATTCGCGAGCCATGCGAGAGAAATAGCGACTTAAACTTCCAGTATCA 152  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 326 lleaAlalYAlglYlYleAsPrGlyArGsrleuValThrlyAsnserPheArGlyrleu 345  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 153 ACGACTATTTCAGAAAGTTTTCGACCTGAAACGAGATACAGCTTCTTTGCTGAA 212  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 346 HlsThrleuileAsnleuGlYserlArPrGluProAsnMetThrValleuTrpserGlu 365  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 213 AATGTTGAAAGAGAGCTTATATTATTAGTGAACCTGACTGCTGAAATGTA 272  
|||||:::|||||:::|||||:::|||||:::|||||  
Db 366 Asnleu---ProGlulserPheYleYserPheCysAlaGlulmeSerlleuThrAsper 384  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 273 TCCGAGTA-----ACGCTTACCTGTTTATGATGAGATATGCTTA--- 317  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 385 llegInlYrGlulAsnArPrGluIleArPrGluIleArPrGluIleArPrGluIleAla 404  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 318 -----GAGCACTAGTAAT 332  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 405 CysCysValserAlameArGValglYlYsArPheGlnPhePheglYAlaArGysAsn 424  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 333 ACCGCTGATGCTCCCAATCAATTAATCTACAGATTGTGAGAGAAAGGAAATG 392  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 425 leuAlalYsCysleuLeuAlalaleAsnGlYglYAlaPrGluYs----- 440  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 393 TTAAAGATCATGTGCGAGGTTACCTAAT---CAACAATTAATCTTTAAAGCGCTG 449  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 441 -----LysglYlYleYValValPrAsPrleGlulPrGluPrleThrAsPrGluValleu 457  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 450 AATTTACTACCGGGGAAAAAATATCTCTGGAATATTAAGAT---CAGTTATG 506  
:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 458 AsPrYr-----GlulYsVallyGlulAsnTrpPheYsValleuGlulYrweC 473  
:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 507 GTTGTTACTATGTAAC 524  
|||||:::|||||:::|||||:::|||||:::|||||  
Db 474 AlaglyLeuYrValAsn 479  
|||||:::|||||:::|||||:::|||||:::|||||

RESULT 3  
US-09-072-433-30  
Sequence 30, Application US/09072433  
Patent No. 6551814  
GENERAL INFORMATION:  
APPLICANT: Coschigano, Peter W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,433



ORGANISM: Enterococcus faecalis  
US-09-134-000C-6467

Alignment Scores:  
Pred. No.: 0.215 Length: 768  
Score: 81.00 Matches: 43  
Percent Similarity: 35.7% Conservative: 31  
Best Local Similarity: 20.8% Mismatches: 67  
Query Match: 8.6% Indels: 66  
DB: 2 Gaps: 9

US-10-768-093-4 (1-527) x US-09-134-000C-6467 (1-768)

```
OY 66 AAGAAACAAATGCTTTAATTTCTAATTCCTTCCTTCGAGCAGCATGCCAGAACAGAA 125
    |||||  ::::::::::|  |||||  ::
Db 21 LysleuValAlaArgProTyrValLeuMetAlaArgPheGlyLeuGluLeuGln 40
    |||||  ::::::::::|  |||||  ::
OY 126 ATAGCGACTAAATCTCCCGATCAACGACTATTTCAAAAAGTTT 176
    |||||  ::::::::::|  |||||  ::
Db 41 ArgSerThrArgGluGlyLeuLeuAlaThrThrGluHisProThrValPheGlyAsnArg 60
    |||||  ::::::::::|  |||||  ::
OY 177 GCACCTGAACCAAGATACAGCCTCTTTGCTGA 212
    |||||  ::::::::::|  |||||  ::
Db 61 SerTyrHisProTyrTleGlnThrAspPheSerGluThrGlnLeuGluLeuThrPro 80
    |||||  ::::::::::|  |||||  ::
OY 213 -AATGTTGGAAAGAA- -GAGCTTTATTTATTTAGTGAAC 251
    |||||  ::::::::::|  |||||  ::
Db 81 ValAlaAsnSerGlyThrGluMetLeuArgPheLeuAspAlaIleHisAspValAlaArg 100
    |||||  ::::::::::|  |||||  ::
OY 252 TTAATCTGTTCTGAAAT- -GATCCCAAGTAAACGCTACCTGTT 296
    |||||  ::::::::::|  |||||  ::
Db 101 ArgSerIleProGluAspGluMetLeuTrpProLeuSerMetProGlnLeuProThr 120
    |||||  ::::::::::|  |||||  ::
OY 297 TATAGTGAAGATTAGGTTAGAGACTAGTAATAACCGCTAGCTTCCCAATCAATA 356
    |||||  ::::::::::|  |||||  ::
Db 121 LysAspGluGluIleLeuIleAlaLeuLeu -AspGlnTyrAspAlaVal 136
    |||||  ::::::::::|  |||||  ::
OY 357 ATTCACAG- -ATTGTTGATGAGAAAGGAAAAA- - 389
    |||||  ::::::::::|  |||||  ::
Db 137 LeuTyrArgArgTyrLeuAlaLeuGlyTyrGlyLeuArgGlyGlnMetValSerGlyIle 156
    |||||  ::::::::::|  |||||  ::
OY 390 - - - - -AGCTTAAAGATCAGTCGACAG 413
    |||||  ::::::::::|  |||||  ::
Db 157 HisPheAsnPheGluTyrAspGlnAlaLeuIleGlnGlnLeuTyrAspGluGlnSerGlu 176
    |||||  ::::::::::|  |||||  ::
OY 414 GTTACACCTAATCAACAATACTTTTAAAGCG- - - - - 446
    |||||  ::::::::::|  |||||  ::
Db 177 ValThrAspCysLysGlnPheLeuThrLysValItyrMetLysValAlaAlaArgAsnPheLeu 196
    |||||  ::::::::::|  |||||  ::
OY 447 - - - - -CTGAATTATATCTAGCGGGGAAAAAATAATCTCTCGA- - - - - 485
    |||||  ::::::::::|  |||||  ::
Db 197 ArgTyrArgTrpLeuIleThrTyrLeuPheGlyAlaSerProValSerGluAspArgTyr 216
    |||||  ::::::::::|  |||||  ::
OY 486 - - - - -ATATATTAACGATCAG 500
    |||||  ::::::::::|  |||||  ::
Db 217 PheArgValTyrAspAspGln 223
    |||||  ::::::::::|  |||||  ::

RESULT 6
US-09-538-092-314
; Sequence 314, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Manfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15566-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178, 965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
```

SOFTWARE: Curataseqformatcer Version 0.9

SEQ ID NO 314  
LENGTH: 345  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (0)...(0)  
OTHER INFORMATION: Polypeptide Accession Number YGL175C  
US-09-538-092-314

Alignment Scores:  
Pred. No.: 0.501 Length: 345  
Score: 77.00 Matches: 44  
Percent Similarity: 40.6% Conservative: 34  
Best Local Similarity: 22.9% Mismatches: 82  
Query Match: 8.2% Indels: 32  
DB: 2 Gaps: 7

US-10-768-093-4 (1-527) x US-09-538-092-314 (1-345)

```
OY 8 ATTAGAATGTTATGATATACAAATAAATGATTTACCAATTTAATGCTGTATATGA 67
    |||||  |||||  |||||  ::
Db 25 LeuLeuAsnValGlnTyrAspValThrThrLeuIleAlaLysArgValGlnAlaLeuGln 44
    |||||  |||||  |||||  ::
OY 68 GAAAAAATGCTTTAATTTCTAATTCCTTCATTCGCGACCA- - - - - 112
    |||||  |||||  |||||  ::
Db 45 - - - - -AsnArgAsnLysCysValLeuGluGluProAsnSerLysLeuAla 59
    |||||  |||||  |||||  ::
OY 113 - - - - -TGCCAGAACAGAAATAGCGACTAAAACTCCAGTATCAACGACTATTC 163
    |||||  |||||  |||||  ::
Db 60 GluIleLeuCysHisGluLysAsnAlaPro--GlnGlnSerSer-GlnThrSerAlaG1 78
    |||||  |||||  |||||  ::
OY 164 AAAAAGTTTTCACCTGACCGACCAAGATACAGCTTCTTTGCTGAAAATGTTGAAA 223
    |||||  |||||  |||||  ::
Db 78 YProGluGluAsnAspSerGluAspPheIleLeuThrGlnPheAspGluAspIleLysLeu 98
    |||||  |||||  |||||  ::
OY 224 GGAAGAGCTTTATTTATTTAGTGAAC- - - - -TT 253
    |||||  |||||  |||||  ::
Db 98 SgluSerIleGluValHisTyrArgAsnGluAsnLysHisThrValGlnLeuProLeuVal 118
    |||||  |||||  |||||  ::
OY 254 AACGTTCCTGAAAATGATATCCAGGTAAAGGCTACCGCTTTATGATGAAGATTATG 313
    |||||  |||||  |||||  ::
Db 118 IthMetProProAsnArgHisLysArgLysIleSerGluPheSerProLeuAsnG1 138
    |||||  |||||  |||||  ::
OY 314 GTTAGAGCACTAGTAATAACCGCTGATGCTTCCCAATCAATAATCAACAGATTGTA 373
    |||||  |||||  |||||  ::
Db 138 YLeuAsnAsnLeuSerAspLeuGluAspCysSerAspThrValIleHisGluLysAspAs 158
    |||||  |||||  |||||  ::
OY 374 TGAGAAAGGAAAAAATGTTAAAGATCATGTGCGACAGGTTTACACTAATCAACAAT 433
    |||||  |||||  |||||  ::
Db 158 AspLysGluAsnLysThrArgLysLeuLeuGlyIleGluLeu--GluAsnProGluSe 177
    |||||  |||||  |||||  ::
OY 434 AACTTTAAAGCGTGAATTTACTAGCGGGGAAAAAATAATCTCTCGAATAT- - - - - 489
    |||||  |||||  |||||  ::
Db 177 rThrSerProAsnLeu--TyrLysAsn-ValLysAspAsnPheLeuPheAsnThr 196
    |||||  |||||  |||||  ::
OY 490 - - - - -ATACGATCAGGTTATGTTGTT 513
    |||||  |||||  |||||  ::
Db 196 HisAsnProLeuThrLysArgAlaTrpIleLeu 206
    |||||  |||||  |||||  ::

RESULT 7
US-09-252-991A-21015
; Sequence 21015, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

US-10-768-093-4 (1-527) X US-09-543-681A-6955 (1-186)

```

QY      1  TTGACCATTAAGCATGGTATGATTAACAACTAAATATGTTTATGACAAATATTAATGGTGT- 59
      ::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      37  MetThArgGlnSerLeuLeuThrProHnThyGlyThrArgGlyrGsetaArpAlaHnIGlyAsn 56
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      60  -----TATATGAAGAAACAATGGTATTATTCGATCTCTGGTCTCA 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      57  SerGlnGlyAlaGlnGlnIleuGlySerGluIleuGlyTrpIleuValLeu----- 70
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



102 TTCGGCACCATGCGAGAAACAGAAATAGCGACTAAATCTCCAGTATCAAGCATATT 161  
DB 71 -----LysHISGLNArgAspGluAsnGluAsnLeuLysTyrProThrAspPheGluVal 88  
162 TCAAAAATTTTTTTCGACCTGACACGAAATACAGCCTCTTTTGGT----- 209  
DB 89 VALHISGLYpPhePheValGlnSerAspMetSerThrAspAspPheGlyPheAsnHisThr 108  
210 ---GAAAATGTTGGAAGGAA-----GGAGCTTTATATTATTAGTGTG 248  
DB 109 PheGlnAsnMetGlyArgLysSerThrGluAspIleLysValGluLeuIleGlnLeu 128  
249 AACTTAATGTTCTCCGAAAATGATATCCGAGTACGGTCTACCTGTTATGATGAAGT 308  
DB 129 AsnAsnGluValAspAspAsnValGlnTyr----- 138  
309 TATGGGTAGAGCGATGTAATACCGGTATGCTTCCCATCAATATATACAGATT 368  
DB 139 -----LysLeuIlePheLeuValArgHisGlyHisSerThrGlnThrGlnIle 154  
369 GTTGATGAGAAAGGAGAAAATGTTAAAGATGATCGTGCAGAGTTACACCTAATCA 428  
DB 155 ValGluLeuGluGlyLeuHisLeuThrPheHisGlyLysAsnCysLysLeuThrGlyAsnAsp 174  
429 CAATTAATCTTT 440  
DB 175 GlnPheThrTyr 178  
RESULT 10  
US-09-134-001C-4037  
Sequence 4037, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4037  
LENGTH: 1076  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4037  
Alignment Scores:  
Pred. No.: 0.993 Length: 1076  
Score: 76.50 Matches: 41  
Percent Similarity: 40.4% Conservative: 24  
Best Local Similarity: 25.5% Mismatches: 79  
Query Match: 8.1% Indels: 17  
Gaps: 7  
US-10-768-093-4 (1-527) x US-09-134-001C-4037 (1-1076)  
DB 60 TATATGAGAAAACAACTTGGTTAATCTTAATTTCT---GCTTCATTCGGCAGCATGCC 116  
DB 570 TTTATGTLGValThrTyrLysAspIleValIleLeuGlnArgSerPheGlyGlnAlaArg 589  
DB 117 AGACAGAAATGCGACATAAAGC-----TTCGCAATGATCAAGCACTATTTCAAAAAGT 170  
DB 590 AsnLeuGlnGlnAlaPheLysAsnAsnAspIleProPheHisValAsnSerLysGluGly 609  
DB 121 TTTT---GCACTGACACCAAGATACAGCTTTCTTTGGTGAATGTTGGAAGGAA 227  
DB 610 TyrPheGluGlnThrGlyValArgLeuValLeuSerPheLeuArgHisIleAspAsnPro 629  
DB 228 GAGCTTTATTTATAGTGTGAACCTTAATCTGTTCCGAAAATGTATCCAGGTACGGTC 287

DB 630 LeuGlnAspIleTyrLysValGlyLeu-----MetArgSerValIle 643  
DB 288 TACCCTGTTTATGATGAAATGATGAGTTAGACGACTAGTAATACCGCTGAT----- 341  
DB 644 TyrGlnPheThrGlnGluGluGluValAlaLysIleArgValValSerProHisAspAspTyr 663  
DB 342 GCTTCCCAATGATATATCTACGAAATGTTGATGAGAAAGGAAAAATTTAAAGAT 401  
DB 664 PheTyrGlnSerIleLysAsnTyrMetIleAspGlyLysValAspSerArgLeuValAsp 683  
DB 402 CATGTCGACAGGTTATACCACTATCAAAATACTTTAAAGCGCTGATTAATATAGTGC 461  
DB 684 LysLeuAsnArgPheIleGlnAspIleGln-----LysTyrGlnAsnTyrSerGln 700  
DB 462 GCGGAAAAA-----AAATATCTCTGGAATATATATACGATACAGTTATGCTGTAC 515  
DB 701 SerGlnProValTyrGlnLeuIleAspLysPheTyrAsnAspHisPheValIleGlnTyr 720  
DB 516 TAT 518  
DB 721 Phe 721  
RESULT 11  
US-09-328-352-6886  
Sequence 6886, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 6886  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-6886  
Alignment Scores:  
Pred. No.: 0.547 Length: 218  
Score: 76.00 Matches: 42  
Percent Similarity: 37.9% Conservative: 24  
Best Local Similarity: 24.1% Mismatches: 80  
Query Match: 8.1% Indels: 28  
Gaps: 7  
US-10-768-093-4 (1-527) x US-09-328-352-6886 (1-218)  
DB 60 TATATGAGAAAACAACTT-----GCTTAAATCTTAATTTCT 95  
DB 40 PheMetLysLysPheIlePheIleAlaLysSerLeuPheAsnIleAlaAlaGln 59  
DB 96 GCTTCATTCGCGACCGACATCCAGAAACAGAAATAGCACTAAATCTCCAGTATCAAG 155  
DB 60 AlaAlaAspGlyThrIleThrIleAsnGlyLeuValThrAspAsnThrCysThrIleAsp 79  
DB 156 ACTATTTCAAAAAGTTTTCGACCTGACACGAAATACAGCACTTCTTTTGGTGAAT 215  
DB 80 ThrGlyAspLysAsnLeuThrValAsnLeuProThrValSerSerGlnSerLeuLysAsn 99  
DB 216 GTTGGAAGC---GAAAGCTTTATATTATTTAGTGAATTAATCTGTTCCGAAAATGTA 272  
DB 100 AlaGlyAspValAlaGlyArgThrProPheGlnIleAsnLeuThrAsnCysAlaSerVal 119  
DB 273 TCCAGAGTAAAGGTCTAC-----CTGTTATGATGAAAGATTAGGTTAGSAGACTA 326  
DB 120 GlyLysValAlaThrTyrPheGlnProGlyAlaThrValAspPheAsnThrGlyArgLeu 139  
DB 327 GTAAT-----ACCGCTGATGCTTCCCAATCAATTAATCTACAGATT----- 368



```
DB      140 LeuaSnglnAlnIthrSerGlyAlaIleAlaAenValaEniIegInLeuLgylSerAsn 159
        |||
QY      369 -----GTTGATGAGAAAGGGAAAAAATTGTAAAAGTCAAGCGCAGAG 413
        |||
DB      160 AsmAlaValIleProValIleAlaValSerGlyAlaGlnAspAsnSerGlnTrp 179
        |||
QY      414 GTTACA-----CCTAATCAACAAATTACTTTAAAGCGCTGAATTACTAGC 461
        |||
DB      180 ValAenValSerAlaGlyAsnAlaAspLeuAenValFYrAlaGluTYrAlaIthr 199
        |||
QY      462 GGGGAAAAAAAAATAATCTCTCGAATATATTAACATCAGTCAGTT 503
        |||
DB      200 GlyAlaSer-----ThraIacIyThrValIthrSerGlnVal 211
        |||
```

RESULT 12  
US-09-252-991A-23709  
Sequence 23709, Application US/09252991A

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OR INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ. ID NOS: 33142  
 SEQ. ID NO 23709  
 LENGTH: 527  
 TYPE: PSEQ

ORGANISM: *Pseudomonas aeruginosa*  
US-09-252-991A-23709

Alignment scores:	
Prid. No.:	0.827
Length:	52
Score:	76.00
Percent Similarity:	44.8%
Beet Local Similarity:	29.1%
Query Match:	8.1%
DB:	2
Length:	52
Matches:	27
Conservative:	16
Mismatches:	37
Indels:	16
Gaps:	5

US-10-768-093-4 (1-527) X US-09-252-991A-23709 (1-527)

Qy	76	GGTTAAATCTAATTTGCTCATTC	-----GGCAGC	110
		:    :    :		
Db	129	GlyValLeuLeuValMetAlaArgPheArgAggIleLeuGluIleAspProLeuGlyArg	148	
Qy	111	CATCCGAGAACAGAAATATAGCGATCAAAAATTCCAGATACAGCATATTTCAAAAGT	170	
Db	149	PheAlaArgValGlnProGlyValArgAsnLeuAlaIleSerGlnAlaAlaAlaProHis	168	
		:    :    :		
Qy	171	-----TTTTTGACCTGAACCA-----GCAATACAGCCTCTTTGGGAAATGTT	218	
Db	169	GlyLeuTyTyAlaProAspProSerSerGlnIleAlaCysSerIleGlyAlaAsnAla	188	
		:    :    :		
Qy	219	GGAAAGAA-----GGAGCTTATTATTATAGGTGAACCTTAACGTCTCGAAAAATGTA	272	
		:    :    :		
Db	189	AlaGluAsnAlaLeuValValHisCysValLeuValTyrGlyLeuThrVal-----HisAsnLeu	207	
Qy	273	TCCAGAGTAACGCTCAACCTCTTTATATGAAGATATAGGTTAGGA	320	
Db	208	LeuGlnValAspIleValThrLeuGluGlyGlnArgLeuSerLeuGly	223	
		:    :    :		

RESULT 3  
US-09-902-540-11397  
Sequence 11397, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.

```

: APPLICANT: Slater, Steven C.
: APPLICANT: Miegand, Roger C.
: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(15849)B
: CURRENT APPLICATION NUMBER: US/09/902,540
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: 60/217,883
: PRIOR FILING DATE: 2000-07-10
: NUMBER OF SEQ ID NOS: 16825
: SEQ ID NO 11397
: LENGTH: 603
: TYPE: PR1
: ORGANISM: Myxococcus xanthus
US-09-902-540-11397

```

ORGANISM: *Myxococcus xanthus*  
US-09-902-540-11397

Alignment Scores:	
Pred. No.:	length: 603
Score:	Matches: 75.50
Percent Similarity:	Conservative: 41.6%
Best local Similarity:	Mismatches: 26.5%
Query Match:	Indels: 8.0%
DB:	Gaps: 2
	Coverage: 6

US-10-768-093-A (1-527) X US-09-902-540-11397 (1-603)

```

Oy      141 TTCCCAAGATATCAAG-----ACTATT 161
      |||||
Db      304 PNEFIVAlaepSerAlaepTyrGluuAenLeuAryaAlaLeuAlaValLeu 323
      |||||
Oy      162 TCMAAAAGTTTGTTCACCTGAACCAAGAAATACAGCCTTCT-----TTTGGTGAAT 215
      ||| |||
Db      324 AaaAepSerAlaPheThrTyrGluuProGluSerSerThrAlaLeuGlyPheGlyPheArg 343
      ||| |||
Oy      216 GTTGGAAAGAAAGAGCTTTATATTATTAAGTGAACTTAACTGTTCCGTGAATGTAATCC 275
      ||| |||
Db      344 CyGlyTyrLeuGlyLeuLeuHisMetGluLe-----ValGlnGluArgLeu-- 359
      ||| |||
Oy      276 CAGGTACAGGTCTACCCCTTTATGATGAAATATGGGTTAGAGACAGTACGTAAATATACC 335
      ||| |||
Db      360 -----GluuArgGluuTyrAsnLeu--AaaLeuGlyLeuThr 370
      ||| |||
Oy      336 GCTGATGCTTCCCAATCATATATCTACACAGATTGTGTGATGAAAGGAAAAAATGTTA 395
      ||| |||
Db      371 Ala-----ProSerValValTyrArgLeuThrSerTyrGlyGluThrLeu 387
      ||| |||
Oy      396 AAAGATCATGTGCGAGAGTTACACTTAATCAACAATA 434
      ||| |||
Db      388 ValAaaPaaPheAlaLeuLeuProGluValGlnHisIle 400
      ||| |||

```

RESULT 14  
US-09-541-782-2  
Sequence 2, Application US/09541782

```

1  GENERAL INFORMATION:
2  APPLICANT: Nislow, Corey
3  APPLICANT: Sakowicz, Roman
4  APPLICANT: Berard, Christophe
5  TITLE OF INVENTION: Antifungal Assay
6  FILE REFERENCE: 1015
7  CURRENT APPLICATION NUMBER: US/09/541,782
8  CURRENT FILING DATE: 2000-04-03
9  NUMBER OF SEQ ID NOS: 10
10 SOFTWARE: FastSeq for Windows Version 4.0
11 SEQ ID NO 2
12 LENGTH: 1184
13 TYPE: PRT
14 ORGANISM: A. nidulans
15 US-09-541-782-2

```

Alignment Scores:		
Pred. No.:	1.41	length:
Score:	75.50	Matches:
Percent Similarity:	38.6%	Conservative:
		23

Best Local Similarity: 24.7% Mismatches: 57  
 Query Match: 8.0% Indels: 45  
 DB: 2 Gaps: 8

US-10-768-093-4 (1-527) x US-09-541-782-2 (1-1184)

```

Oy 72 ACAATGTTTAATTCATATCTTCTTCATTCGGACCATGCGAAGAAATACCG 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ThrIaGlyIleLeuLeu-----GInGInGlySerHisLeuValAlaIa 282
Oy 132 ACTAA-----AACTCCAGATACAGACTTT 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 ThrIyScySaenbPLeuSerSerArgSerHisThrValPheThrIleThrValAsnIle 302
Oy 162 TCAGAAAGTTTTTTTGACCTGAAACAGATACAGCT----- 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 LysArgThrThrGInuSerGlyGInuIuTyValCysProGlyLyLeuAsnLeuValasp 322
Oy 201 ---TCCTTGGTGAATGTGTGAAAGAGAGCTTTATTATTATTTAGTGAACTTA 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 LeuIaGlySerGInuAsnIleGlyArgSerGlyAla----- 334
Oy 258 GTTCTGAAATGTATCCAGCTACGCTCTACCTGTTATGATGAATATGAGTTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 -----GInuAsnLySaArgAlaThrGInuIaGlyLeuIleAsnLySerLeuLeuThrLeu 352
Oy 318 GACGACTAGTAATATAC---GCTGATGCTTCCCAATCAATATCAATGATGTTGAT 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 GlyArgValIleAsnAlaLeuValAspLySerGInHisIleProTyArg-----Glu 370
Oy 375 GAGAAAGGAAAAAATGTTAAAGATCAT-----GCT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 SerIySeuThrArgLeuGInuAspSerLeuGlyArgThrLySerHisIleIle 390
Oy 408 GCAGAGTTTACACT-----AATCAACAATAATCTTTAAAGCGCTGAATTACT 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 AlaThrMetSerProAlaArgSerAsnLeuGInuIuThrIleSerThrLeuAspTyAla 410
Oy 459 AGCGGGAAAAAATA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 PheArgAlaLySaenIle 416
  
```

RESULT 15  
 US-09-723-820-2

; Sequence 2, Application US/09723820

; Patent No. 6468760

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

; TITLE OF INVENTION: Antifungal Assay

; FILE REFERENCE: 1015

; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: 09/541,782

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1184

; TYPE: PRT

; ORGANISM: A. nidulans

US-09-723-820-2

Alignment Scores:

Pred. No.: 1.41

Score: 75.50

Percent Similarity: 38.6%

Best Local Similarity: 24.7%

Query Match: 8.0%

Length: 1184

Matches: 41

Conservative: 23

Mismatches: 57

Indels: 45

Gaps: 8

```

Oy 72 ACAATGTTTAATTCATATCTTCTTCATTCGGACCATGCGAAGAAATACCG 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ThrIaGlyIleLeuLeu-----GInGInGlySerHisLeuValAlaIa 282
Oy 132 ACTAA-----AACTCCAGATACAGACTTT 161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 ThrIyScySaenbPLeuSerSerArgSerHisThrValPheThrIleThrValAsnIle 302
Oy 162 TCAGAAAGTTTTTTTGACCTGAAACAGATACAGCT----- 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 LysArgThrThrGInuSerGlyGInuIuTyValCysProGlyLyLeuAsnLeuValasp 322
Oy 201 ---TCCTTGGTGAATGTGTGAAAGAGAGCTTTATTATTATTTAGTGAACTTA 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 LeuIaGlySerGInuAsnIleGlyArgSerGlyAla----- 334
Oy 258 GTTCTGAAATGTATCCAGCTACGCTCTACCTGTTATGATGAATATGAGTTA 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 -----GInuAsnLySaArgAlaThrGInuIaGlyLeuIleAsnLySerLeuLeuThrLeu 352
Oy 318 GACGACTAGTAATATAC---GCTGATGCTTCCCAATCAATATCAATGATGTTGAT 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 GlyArgValIleAsnAlaLeuValAspLySerGInHisIleProTyArg-----Glu 370
Oy 375 GAGAAAGGAAAAAATGTTAAAGATCAT-----GCT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 SerIySeuThrArgLeuGInuAspSerLeuGlyArgThrLySerHisIleIle 390
Oy 408 GCAGAGTTTACACT-----AATCAACAATAATCTTTAAAGCGCTGAATTACT 458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 AlaThrMetSerProAlaArgSerAsnLeuGInuIuThrIleSerThrLeuAspTyAla 410
Oy 459 AGCGGGAAAAAATA 476
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 411 PheArgAlaLySaenIle 416
  
```

Search completed: April 10, 2006, 10:08:36  
 Job time : 38.5 secs

US-10-768-093-4 (1-527) x US-09-723-820-2 (1-1184)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 10, 2006, 10:07:45 ; Search time 25.1 Seconds  
(without alignments)  
1754.550 Million cell updates/sec

Title: US-10-768-093-4  
Sequence: 1 TTGCACATTAAGATGTTA.....TTGCTACTATGTAACATA 527

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1867569 seqs, 417829326 residues  
Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=eframe+n2p model -DEV=xlp  
-Q=/abses/ABSSWB\_topool/US10768093/runat\_10042006\_09309\_4677/app\_query.fasta\_1  
-DB=Published\_Applications\_AA\_Main -QFMT=fastan -SUFFIX=n2p.rapbm  
-MIMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1  
-MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext  
-HEAFSIZE=500 -MILEN=0 -MAXLEN=200000000 -HOST=aabs02p  
-USRR=US10768093 @CN 1.1 307 @runat\_10042006\_09309\_4677 -NCPU=6 -ICPU=3  
-NO MMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FAGAPOP=6 -FAGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	785	83.6	154	4	US-10-768-093-5
2	785	83.6	154	4	US-10-754-641-5
3	699	74.4	136	4	US-10-768-093-9
4	699	74.4	136	4	US-10-754-641-9
5	118	12.6	167	4	US-10-768-093-6
6	118	12.6	167	4	US-10-754-641-6
7	113.5	12.1	146	4	US-10-768-093-10
8	87.5	9.3	911	4	US-10-197-666A-100
9	85	9.1	541	4	US-10-282-122A-77158
10	84	8.9	439	4	US-10-365-493-2489
11	84	8.9	740	4	US-10-267-989-25

12	84	8.9	740	4	US-10-357-567-25	Sequence 25, Appl
13	83.5	8.9	157	4	US-10-424-599-190153	Sequence 190153,
14	83	8.8	249	4	US-10-424-599-250712	Sequence 250712,
15	83	8.8	750	5	US-10-501-282-3988	Sequence 3988, Ap
16	82	8.8	2094	4	US-10-437-963-161607	Sequence 161607,
17	79.5	8.5	429	4	US-10-359-493-17433	Sequence 17433, A
18	79.5	8.5	951	4	US-10-197-666A-94	Sequence 94, Appl
19	79.5	8.5	1171	4	US-10-197-666A-96	Sequence 96, Appl
20	79.5	8.5	1171	4	US-10-311-034-8	Sequence 8, Appl
21	78.5	8.4	1198	4	US-10-001-215-2	Sequence 2, Appl
22	78.5	8.4	1198	4	US-10-197-666A-98	Sequence 98, Appl
23	77.5	8.3	357	4	US-10-282-122A-656978	Sequence 65978, A
24	77	8.2	458	4	US-10-425-115-223762	Sequence 223762,
25	77	8.2	471	3	US-09-815-242-11936	Sequence 11936, A
26	76.5	8.1	330	5	US-09-815-242-11936	Sequence 7096, Ap
27	76.5	8.1	442	4	US-10-452-024-76	Sequence 76, Appl
28	76.5	8.1	1076	4	US-10-724-972A-5848	Sequence 5848, Ap
29	76	8.1	348	4	US-10-437-963-136628	Sequence 136628,
30	76	8.1	499	4	US-10-389-647-664	Sequence 664, App
31	76	8.1	557	4	US-10-369-493-21384	Sequence 21384, A
32	76	8.1	2358	4	US-10-282-122A-68607	Sequence 68607, A
33	75.5	8.0	246	4	US-10-282-122A-67315	Sequence 67315, A
34	75.5	8.0	1218	5	US-10-470-048B-444	Sequence 444, App
35	75.5	8.0	1224	4	US-10-282-122A-70506	Sequence 70506, A
36	75	8.0	640	4	US-10-282-122A-44691	Sequence 44691, A
37	75	8.0	736	3	US-09-815-242-11253	Sequence 11293, A
38	75	8.0	736	4	US-10-282-122A-58624	Sequence 58624, A
39	74.5	7.9	401	4	US-10-282-122A-54471	Sequence 54471, A
40	74.5	7.9	430	4	US-10-282-122A-60954	Sequence 60954, A
41	74.5	7.9	497	4	US-10-724-972A-6030	Sequence 6030, Ap
42	74.5	7.9	499	4	US-10-369-493-23554	Sequence 23554, A
43	74.5	7.9	581	5	US-10-450-763-48573	Sequence 48573, A
44	74.5	7.9	2849	4	US-10-437-963-151131	Sequence 151131,
45	74	7.9	1666	6	US-11-097-143-42954	Sequence 42954, A

## ALIGNMENTS

RESULT 1  
US-10-768-093-5  
Sequence 5, Application US/10768093  
Publication NO. US20040156828A1  
GENERAL INFORMATION:  
APPLICANT: Wolf, Marcia K  
Cassels, Fred J  
Boedeker, Edgar C  
TITLE OF INVENTION: Transformed Bacteria Producing GS6  
Antigens as Vaccines  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hendricks and Assoc  
STREET: P.O. Box 2509  
CITY: Fairfax  
STATE: VA  
COUNTRY: US  
ZIP: 22031  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/768, 093  
FILING DATE: 02-Feb-2004  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/479, 877B  
FILING DATE: 10-Jan-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna M  
REGISTRATION NUMBER: 32,535  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703/425-8405

**Tue Apr 11 08:36:00 2006**

us-10-768-093-4.n2p.rapbm

Page 2

```

TELEFAX: 703/425-8406
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
    LENGTH: 154 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: unknown
    MOLECULE TYPE: peptide
    HYPOTHEICAL: NO
    ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-768-093-5

Alignment Scores:
Pred. No.:      Length:      154
Score:          Matches:     154
Percent Similarity: 785.00
Best Local Similarity: 100.0%
Query Match:    100.0%
DB:             Gaps:        0
               Indels:       0
               Mismatches:   0
               Deletions:    0
               Insertions:   0
               Gaps:         0
               Indels:       0
               Mismatches:   0
               Deletions:    0
               Insertions:   0

US-10-768-093-4 (1-527) x US-10-768-093-5 (1-154)
QY      63  ATGAGAAAAACAATTGGTTAATTCCTTCAATTCGGACGCATCCAGAACAA 122
           |
Db      1  MetVSLyVethrIIeGIleuLeulleuLeulaIasSerPheglYseHIsalaIarYthr 20
QY      123 GAATATGCACATAAAAATTCCAGATCAACGAATATTTCAAAGAATTTTTTGACACT 182
           |
Db      21  GIUIIEALITThLyehsnpherovalserThrIIIeserlyserPhenhalPro 40
QY      183 GAACACGAATACAGCCTTCTTTGGTGAAAATGTGGAAAAGAGAGCTTTATTT 242
           |
Db      41  GIUPErArGIeGINroSerPheglYguaenValIGlYlYsgLUGlYAlaleulephe 60
QY      243 AGTGCGAATCTTAACGTCTTCGAAAATGATCCAGGTACCGGTACCCTGTTATGAT 302
           |
Db      61  SetValAsnleuthIValProglumenValserGlinalmValIYProvalYTrasp 80
QY      303 GAAGATTATGGGTTAGACGACTAGTAATAACCGGTGATGCTCCCAATCAATATCTAC 362
           |
Db      81  GIUAsPTYGIYleuclYargleuValaenthrrlaIaspaIasergInserlleIetyr 100
QY      363 CAGATTGTTGATGAGAAAGGAAAAAAAGTTAAAAAGATCATGTGCGAGAGTTACACT 422
           |
Db      101  GlnIleValAspGIulYsgLIySLyswctleuLYAspHIsGLYAlaGIuValThPro 120
QY      423 AATCAACAATATACTTTTAAAGCGCTGAATTTACTAGCGGGGAAAAAAATATCTCT 482
           |
Db      121  AengInGlnIIIEthrhelysalaleuanrYrThserSeIGlulYlYslIeserPro 140
QY      483 GGATATATATACGATCAGGTTATGTTGGTTACTATGTAAAC 524
           |
Db      141  GIYIIEtYrAsnaAspGIuValmetValGIYTYrYValasn 154

RESULT 2
US-10-754-641-5
Sequence 5, Application US/10754641
Publication No. US20050025787A1
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research and Materiel Command
APPLICANT: Wolf, Marcia K.
APPLICANT: Casella, Frederick J.
APPLICANT: Boedeker, Edgar C.
TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES
FILE REFERENCE: 034047, 033,3
CURRENT APPLICATION NUMBER: US/10/754,641
PRIOR FILING DATE: 2004-01-12
PRIOR APPLICATION NUMBER: 09/479,877
PRIOR FILING DATE: 2000-01-10
PRIOR APPLICATION NUMBER: 08/788,145
PRIOR FILING DATE: 1997-01-24
PRIOR APPLICATION NUMBER: 08/243,482
```

```

      / PRIOR FILING DATE: 1994-05-13
      / NUMBER OF SEQ ID NOS: 9
      / SOFTWARE: PatentIn version 3.2
      / SEQ ID NO 5
      / LENGTH: 154
      / TYPE: PRF
      / ORGANISM: Escherichia coli
      / US-10-754-641-5

Alignment Scores:
Pred. No.:      1,09e-82      Length:      154
Score:          785.00      Matches:      154
Percent Similarity: 100.0%      Conservative: 0
Best Local Similarity: 100.0%      Mismatches: 0
Query Match:      83.6%      Indels:      0
DB:              5      Gaps:      0

US-10-768-093-4 (1-527) x US-10-754-641-5 (1-154)

QY      ATGAGAGAAACCAATTGGTTTAATTCTTAATCTTGCTTCATTCCGACGCATGCCAGACA 122
Db      1 MetLyAsrThrIleGolyLeuIleLeuIleLeuIleAsrPheGlySerHsAlaAlyThr 20

QY      GAATATGCGCATTAAGAACTCCCACTATCAACGACTATTTCAAAAAGTTTTTTTGACACT 182
Db      123 GAATATGCGCATTAAGAACTCCCACTATCAACGACTATTTCAAAAAGTTTTTTTGACACT 182
      21 GlnIleIleThrLysAsnNheProValSerThrThrIleSerLysSerPhePheAlaPro 40

QY      GAACACAGATACAGGCTTCTTTTGGTCAAAAATGTTGGAAAAGAGAGACTTATTTA 242
Db      183 GAACACAGATACAGGCTTCTTTTGGTCAAAAATGTTGGAAAAGAGAGACTTATTTA 242
      41 GluProArgIleGlnProSerPheGlyGlnAsnValGlyLysGlnGlyAlaLeuLeuPhe 60

QY      AGGTGTGACTTAATCTGTTCCCTGAAAATGTATCCAGGTAAAGGCTCCCTGTTATGAT 302
Db      243 AGGTGTGACTTAATCTGTTCCCTGAAAATGTATCCAGGTAAAGGCTCCCTGTTATGAT 302
      61 SerValAsnLeuThrValProGlnAsnValSerGlnValThrValTYrProValTYrAsp 80

QY      GAAGATTATGAGGTTAGAGACGACTAGTAATAATACCGCTGATGCTTCCCAATCATATCTAC 362
Db      303 GAAGATTATGAGGTTAGAGACGACTAGTAATAATACCGCTGATGCTTCCCAATCATATCTAC 362
      81 GluAspTYrGlyLeuGlyLysIleuValAsnThrAlaAspAlaSerGlnSerIleIleTYr 100

QY      CAGATTGTTGATGAGAAAGGAAAAAAATGTTAAAGATCATGTCGACAGGTTACACCT 422
Db      363 CAGATTGTTGATGAGAAAGGAAAAAAATGTTAAAGATCATGTCGACAGGTTACACCT 422
      101 GlnIleValAlaSpGlnIleGolyLysLysMetLeuLysAspHisGlyAlaGlnValThrPro 120

QY      AATCAACAATAATACCTTTTAAAGCGCTGAATATATACGTAGCGGGAAAAAAATATCTCCT 482
Db      423 AATCAACAATAATACCTTTTAAAGCGCTGAATATATACGTAGCGGGAAAAAAATATCTCCT 482
      121 AsnGlnGlnIleThrPheLysAlaLeuValenTYrThrSerGlyGlnLysIleSerPro 140

QY      GGAATATATACGATCAGGTTATGAGTTAGTTACTATGTAATAC 524
Db      483 GGAATATATACGATCAGGTTATGAGTTAGTTACTATGTAATAC 524
      141 GlyIleTYrAsnAspGlnValIleValGlyTYrTYrValAsn 154

RESULT 3
US-10-768-093-9
/ Sequence 9, Application US/10768093
/ Publication No. US20040156829A1
/ GENERAL INFORMATION:
/ APPLICANT: Wolf, Marcia K
/ Casseis, Fred J
/ TITLE OF INVENTION: Transformed Bacteria Producing GSE
/ Antigenes as Vaccines
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Hendricks and Assoc
/ STREET: P.O. Box 2509
/ CITY: Fairfax
/ STATE: VA
/ COUNTRY: US
/ ZIP: 22031
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/768,093  
FILING DATE: 02-Feb-2004  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,877B  
FILING DATE: 10-Jan-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna M  
REGISTRATION NUMBER: 32,535  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703/425-8405  
TELEFAX: 703/425-8406  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 136 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-10-768-093-9

Alignment Scores:  
Pred. No.: 1,23e-72 Length: 136  
Score: 699.00 Matches: 136  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 74.4% Indels: 0  
DB: Gaps: 0

US-10-768-093-4 (1-527) x US-10-768-093-9 (1-136)

QY 117 AGAAGCAATATAGGACTATAAACTTCCAGTATCAACGACTATTTCAGAAAAGTTT 176  
Db 1 ArgThcIuIleAtrhLyAsnPheProValSerThrThrIleSerLySerPhe 20

QY 177 GCACCTGAACCAAGATACAGCCTCTTTGGTGAATAATGTGAAAGAGAGAGCTTTA 236  
Db 21 AlapProgluProlargIleGlnProSerPheGlyGluAsnValGlyLybGluGlyAlaLeu 40

QY 237 TTAATTAGTGAACCTTAAGTCTTCTGAAAATGTATCCAGTAAAGGCTTACCTGTT 296  
Db 41 LeuphSerValenLeuThrValProgluAsnValSerGlnValThrValTyProVal 60

QY 297 TATGATGAAGATTATGGGTTAGAGCACTAGTAAATACCGCTGATGCTTCCCAATCAATA 356  
Db 61 TyAspGluAspTyrgIyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIle 80

QY 357 ATTCACGAGATTGTGTGAGAAAGGAAAAAATGTTAAAGATCATGTGTCAGAGGTT 416  
Db 81 IleTyrcIlnIleValAspGluLybGlybLybMetLeuLybAspHisGlyAlaGluVal 100

QY 417 ACACTATCAACAATAACTTTTAAAGCGCTGAATTAATACTAGCGGGGAAAAAATA 476  
Db 101 ThrProAsnGlnGlnIleThrPheLybAlaLeuAsnTyThrSerGlyGluLybIle 120

QY 477 TCTCCGGAATATATACGATCAGGTTATGGTGTACTATGTAAC 524  
Db 121 SerProGlyIleTyraAsnAspGlnValMetValGlyTyTyraValAsn 136

RESULT 4  
US-10-754-641-9  
Sequence 9, Application US/10754641  
Publication No. US20050025787A1  
GENERAL INFORMATION:  
APPLICANT: U.S. Army Medical Research and Materiel Command  
APPLICANT: Wolf, Marcia K.  
APPLICANT: Casseels, Frederick J.  
APPLICANT: Boedeker, Edgar C.

TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES  
FILE REFERENCE: 034047.033.3  
CURRENT APPLICATION NUMBER: US/10/754,641  
CURRENT FILING DATE: 2004-01-12  
PRIOR APPLICATION NUMBER: 09/479,877  
PRIOR FILING DATE: 2000-01-10  
PRIOR APPLICATION NUMBER: 08/788,145  
PRIOR FILING DATE: 1997-01-24  
PRIOR APPLICATION NUMBER: 08/243,482  
PRIOR FILING DATE: 1994-05-13  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patent in Version 3.2  
SEQ ID NO 9  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-754-641-9

Alignment Scores:  
Pred. No.: 1,23e-72 Length: 136  
Score: 699.00 Matches: 136  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 74.4% Indels: 0  
DB: Gaps: 0

US-10-768-093-4 (1-527) x US-10-754-641-9 (1-136)

QY 117 AGAAGCAATATAGGACTATAAACTTCCAGTATCAACGACTATTTCAGAAAAGTTT 176  
Db 1 ArgThcIuIleAtrhLyAsnPheProValSerThrThrIleSerLySerPhe 20

QY 177 GCACCTGAACCAAGATACAGCCTCTTTGGTGAATAATGTGAAAGAGAGAGCTTTA 236  
Db 21 AlapProgluProlargIleGlnProSerPheGlyGluAsnValGlyLybGluGlyAlaLeu 40

QY 237 TTAATTAGTGAACCTTAAGTCTTCTGAAAATGTATCCAGTAAAGGCTTACCTGTT 296  
Db 41 LeuphSerValenLeuThrValProgluAsnValSerGlnValThrValTyProVal 60

QY 297 TATGATGAAGATTATGGGTTAGAGCACTAGTAAATACCGCTGATGCTTCCCAATCAATA 356  
Db 61 TyAspGluAspTyrgIyLeuGlyArgLeuValAsnThrAlaAspAlaSerGlnSerIle 80

QY 357 ATTCACGAGATTGTGTGAGAAAGGAAAAAATGTTAAAGATCATGTGTCAGAGGTT 416  
Db 81 IleTyrcIlnIleValAspGluLybGlybLybMetLeuLybAspHisGlyAlaGluVal 100

QY 417 ACACTATCAACAATAACTTTTAAAGCGCTGAATTAATACTAGCGGGGAAAAAATA 476  
Db 101 ThrProAsnGlnGlnIleThrPheLybAlaLeuAsnTyThrSerGlyGluLybIle 120

QY 477 TCTCCGGAATATATACGATCAGGTTATGGTGTACTATGTAAC 524  
Db 121 SerProGlyIleTyraAsnAspGlnValMetValGlyTyTyraValAsn 136

RESULT 5  
US-10-768-093-6  
Sequence 6, Application US/10768093  
Publication No. US20060156829A1  
GENERAL INFORMATION:  
APPLICANT: Wolf, Marcia K  
APPLICANT: Casseels, Fred J  
APPLICANT: Boedeker, Edgar C  
TITLE OF INVENTION: Transformed Bacteria Producing GS6  
Antigens as Vaccines  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Hendricks and Assoc  
STREET: P.O. Box 2509  
CITY: Fairfax  
STATE: VA  
COUNTRY: US

ZIP: 22031  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patientin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/768,093  
FILING DATE: 02-Feb-2004  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,877B  
FILING DATE: 10-Jan-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna M  
REGISTRATION NUMBER: 32,535  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703/425-8405  
TELEFAX: 703/425-8406  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 167 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: peptide  
TOPOLOGY: unknown  
HYPOTHEITICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-768-093-6  
Alignment Scores:  
Pred. No.: 0.000134 Length: 167  
Score: 118.00 Matches: 49  
Percent Similarity: 42.1% Conservative: 26  
Best Local Similarity: 27.5% Mismatches: 67  
Query Match: 12.6% Indels: 36  
Gaps: 9  
DB: 4  
US-10-768-093-4 (1-527) x US-10-768-093-6 (1-167)  
QY 63 ATGAAGAAACATGCTTAAATCTTCTGCTTCCGACCCATGCCAGACA 122  
DB 2 LeuyslysllelleSerallaaleuileala-----GlyThrserylval 19  
QY 123 GAAATAGGACT-----AAAACCTCCAGATCAAGACTATTCAAAAGTTT 173  
DB 20 AsnlaaglyantpGlnTyrlvseryleuaspvalaenvalaenllegluglnasph 39  
QY 174 TTGCACTGAA-----CCAGCAATACAGCTTCTTTGGTGAATGTGGA 221  
DB 40 Ile---ProaplleaspseralavalaargllelleprovalaenTyraaspeaspro 58  
QY 222 AAGGAAGAGCTTTATTTAGTGGAATTAAGTCTTCCGAAATGTATCCAGTA 281  
DB 59 LysleuaspserGlnleuTyrlvalaGlnmetctrlleproalaglyalseralaval 78  
QY 282 ACCGCTACCT-----GTTATGATGAAGATTATGGTTAGAGCACTAGTAAT 332  
DB 79 Lysllealaprothraaspeleuthrserylglnnglnlleglyleuvalaen 98  
QY 333 ACCGCTAGCTTCCCAATCAATACTACAGATTGTGATGAGAAAGGAAAAATG 392  
DB 99 ValaenaspserGlnasmetasentyrlyle-----GAGTTACA 419  
QY 393 TTAAGAATCATGGGCA-----GAGTTACA 419  
DB 111 ArglyaspserGlyalaglyasnphenetalaaglyglnlyserPheProvalyls 130  
QY 420 CCTATCAACAATATCTTTAAAGCGCTGAATTAATACAGCGGGA-----AAA 470  
DB 131 GluamthrserylThrseserallale---TyrlhrglyglnTyrlProasner 149

QY 471 AAAATATCTCGAATAATATACGATCAGTTATGTTGTAATGTAAC 524  
DB 150 GlyTyserserylThryalaglyasnleuthrvalserPheTyserasn 167  
RESULT 6  
US-10-754-641-6  
Sequence 6, Application US/10754641  
Publication No. US20050025787A1  
GENERAL INFORMATION:  
APPLICANT: U.S. Army Medical Research and Materiel Command  
APPLICANT: Wolf, Marcia K.  
APPLICANT: Casseels, Frederick J.  
TITLE OF INVENTION: TRANSFORMED BACTERIA PRODUCING CS6 ANTIGENS AS VACCINES  
FILE REFERENCE: 034047.033.3  
CURRENT APPLICATION NUMBER: US/10/754,641  
PRIORITY FILING DATE: 2004-01-12  
PRIOR APPLICATION NUMBER: 09/479,877  
PRIOR FILING DATE: 2000-01-10  
PRIOR APPLICATION NUMBER: 08/788,145  
PRIOR FILING DATE: 1997-01-24  
PRIOR APPLICATION NUMBER: 08/243,482  
PRIOR FILING DATE: 1994-05-13  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: Patientin version 3.2  
SEQ ID NO: 6  
LENGTH: 167  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-10-754-641-6  
Alignment Scores:  
Pred. No.: 0.000134 Length: 167  
Score: 118.00 Matches: 49  
Percent Similarity: 42.1% Conservative: 26  
Best Local Similarity: 27.5% Mismatches: 67  
Query Match: 12.6% Indels: 36  
Gaps: 9  
DB: 5  
US-10-768-093-4 (1-527) x US-10-754-641-6 (1-167)  
QY 63 ATGAAGAAACATGCTTAAATCTTCTGCTTCCGACCCATGCCAGACA 122  
DB 2 LeuyslysllelleSerallaaleuileala-----GlyThrserylval 19  
QY 123 GAAATAGGACT-----AAAACCTCCAGATCAAGACTATTCAAAAGTTT 173  
DB 20 AsnlaaglyantpGlnTyrlvseryleuaspvalaenvalaenllegluglnasph 39  
QY 174 TTGCACTGAA-----CCAGCAATACAGCTTCTTTGGTGAATGTGGA 221  
DB 40 Ile---ProaplleaspseralavalaargllelleprovalaenTyraaspeaspro 58  
QY 222 AAGGAAGAGCTTTATTTAGTGGAATTAAGTCTTCCGAAATGTATCCAGTA 281  
DB 59 LysleuaspserGlnleuTyrlvalaGlnmetctrlleproalaglyalseralaval 78  
QY 282 ACCGCTACCT-----GTTATGATGAAGATTATGGTTAGAGCACTAGTAAT 332  
DB 79 Lysllealaprothraaspeleuthrserylglnnglnlleglyleuvalaen 98  
QY 333 ACCGCTAGCTTCCCAATCAATACTACAGATTGTGATGAGAAAGGAAAAATG 392  
DB 99 ValaenaspserGlnasmetasentyrlyle-----GAGTTACA 419  
QY 393 TTAAGAATCATGGGCA-----GAGTTACA 419  
DB 111 ArglyaspserGlyalaglyasnphenetalaaglyglnlyserPheProvalyls 130  
QY 420 CCTATCAACAATATCTTTAAAGCGCTGAATTAATACAGCGGGA-----AAA 470  
DB 131 GluamthrserylThrseserallale---TyrlhrglyglnTyrlProasner 149

OY 471 AAATATCTCTGGAATATATACGATCAGCTTATGTTGCTTATGTAAC 524

Db 150 GlyTyrSerSerGlyThrTyrAlaGlyAsnLeuThrValSerPheTyrSerAsn 167

RESULT 7

US-10-768-093-10  
Sequence 10, Application US/10768093  
Publication NO. US20040156829A1  
GENERAL INFORMATION:

APPLICANT: Wolf, Marcia K  
Casele, Fred J  
Boedeker, Edgar C

TITLE OF INVENTION: Transformed Bacteria Producing GSE  
Antigens as Vaccines

NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hendricks and Assoc  
STREET: P.O. Box 2509  
CITY: Fairfax  
STATE: VA  
COUNTRY: US  
ZIP: 22031

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/768,093  
FILING DATE: 02-Feb-2004  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,877B  
FILING DATE: 10-Jan-2000

ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenna M  
REGISTRATION NUMBER: 32,535  
TELEPHONE: 703/425-8405  
TELEFAX: 703/425-8406

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703/425-8405  
TELEFAX: 703/425-8406

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown

MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-768-093-10

Alignment Scores:  
Pred. No.: 0.000436 Length: 146  
Score: 113.50 Matches: 43  
Percent Similarity: 43.0% Conservative: 22  
Best Local Similarity: 28.5% Mismatches: 55  
Query Match: 12.1% Indels: 31  
DB: 4 Gaps: 7

US-10-768-093-4 (1-527) x US-10-768-093-10 (1-146)

OY 135 AAAAATCTCCAGATCAAGACATATTTCAGAAAGTTTTCACCTGAA----- 185  
Db 6 LysSerLeuAlaValIleValIleGluGlnAsnPheLeu-----ProAspIleAspSer 24

OY 186 ---CCACGAATACAGCCTTCTTTGGTGAATAATGTTGAAAGAGAGCCTTATATT 242  
Db 25 AlaValArgIleIleProValIleTyrAspSerAspProLeuAspSerGlnLeuTyr 44

OY 243 AGTGTGAATTAATGTTCTCGAAATGTAATCCCGATGACGCTTACCC----- 293  
Db 45 ThrValGluMetThrIleProIleGlyValIleSerAlaValIleValIleAlaProThrAspSer 64

OY 294 GTTATGATGAAGATTATGGCTTACGACGATGTAATACCGCTGATGCTCCCAATCA 353

Db 65 LeuThrSerSerGlyGlnGlnIleGlyLeuValIleAsnAsnProAspGlnAsn 84

OY 354 ATATATCAACGATTTGTGATGAGAAAGGAAAAAATGTTAAAGATCATGTGCA--- 410

Db 85 MetAsnTyrTyrIle-----ArgLysAspSerGlyAlaGly 96

OY 411 -----GAGTTACACCTTATCAACAATTAATCTTT 440

Db 97 AsnPheMetAlaGlyGlnLysGlySerPheProValIleGluLeuThrSerTyrThrPhe 116

OY 441 AAAGCGTGAATTAATCAACGGGAA-----AAAAAATATCTCGGAATATAT 491

Db 117 SerAlaIle---TyrThrGlyGlyGlnTyrProAsnSerGlyTyrSerSerGlyThrTyr 135

OY 492 AACGATCAGTTATGTTGTTGCTTACTATGTAAC 524

Db 136 AlaGlyAsnLeuThrValSerPheTyrSerAsn 146

RESULT 8

US-10-197-666A-100  
Sequence 100, Application US/10197666A  
Publication NO. US20030092037A1  
GENERAL INFORMATION:

APPLICANT: ASAMI KASEI KABUSIKI KAISYA  
TITLE OF INVENTION: Bkl1 phosphorylation related gene  
FILE REFERENCE: PH-1548US

CURRENT APPLICATION NUMBER: US/10/197,666A  
CURRENT FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: JP 2001-218204  
PRIOR FILING DATE: 2001-07-18

PRIOR APPLICATION NUMBER: JP 2001-263450  
PRIOR FILING DATE: 2001-08-31

PRIOR APPLICATION NUMBER: JP 2002-012176  
PRIOR FILING DATE: 2002-01-21

PRIOR APPLICATION NUMBER: US 60/305,884  
PRIOR FILING DATE: 2001-07-18

PRIOR APPLICATION NUMBER: US 60/316,304  
PRIOR FILING DATE: 2001-09-04

PRIOR APPLICATION NUMBER: US 60/350,027  
PRIOR FILING DATE: 2002-01-23

NUMBER OF SEQ ID NOS: 156  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 100

LENGTH: 911  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-197-666A-100

Alignment Scores:  
Pred. No.: 0.786 Length: 911  
Score: 87.50 Matches: 40  
Percent Similarity: 43.3% Conservative: 25  
Best Local Similarity: 26.7% Mismatches: 62  
Query Match: 9.3% Indels: 23  
DB: 4 Gaps: 8

US-10-768-093-4 (1-527) x US-10-197-666A-100 (1-911)

OY 93 CTGCTTCATTCGGACCCATGCGAGAACGAAATACGACTTAAATCTCCAGATCA 152  
Db 34 MetThrGlyTyrGlySerHisSerLysValTyrSerGlnSerLysValIleProLeuSer 53

OY 153 -----ACGACTATTTCAAAAGTTTTCGACCTGAAACGACGAATACAG--- 197  
Db 54 GlnProAlaThrThrThrValSerThrSerLeuProValProAsnProSerLeuProTyr 73

OY 198 -----CCTTCCTTTGTCGAAATGTTGAAAGAGAGAGCTTATTA 239  
Db 74 GluGlnThrIleValIlePheProGlySerThrGlyHisIleValIleValIleThrSerAlaSerSer 93

QY	240	TTAGGTGAAC-----TTACTGTCCTGAAATGA-----TCCAGGT	281
Db	94	ThSerValThrGlyGlnValLeuGlyGlyProHisLeuLeuMetArgSerThrVal	113
QY	282	ACGGCTACCCCTGTTATGATGAAGATTATGGCTTAGGACGACTAGTAAATACCGCTGAT	341
Db	114	SerLeuLeuAspThrTyr---GlnTyrCysGlyLeuLysArgLysSerGluGluIleGlu	132
QY	342	GCTTCCCATTCATATATCTACAGAGTTGTGATGAGAAAGGAAATAATGTTAAAGAT	401
Db	133	AsnThrSerSerVal-----GlnIleIleGluGluHis---ProPomeIleGlnAsn	149
QY	402	CATGTGACAGAGGTTACACCTTAATCAACAATTAACCTTTAAAGCGCGAAATTATCTACG	461
		:::	:::
Db	150	AsnAlaSerGlyAlaLeuThrValAlaLeuThrAlaThrThrPheThrAlaIleThrSerLysAsnSer	169
QY	462	GGGGAATAAATAATCTCCTGGAATATAT	491
Db	170	GlySerAsn-----SerGluIleAspPylr	177

```

RESULT 9
US-10-282-122A-77158
/ Sequence 77158, Application US/10282122A
/ Publication No. US2004002912A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zykkind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/151,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/267,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 77158
/ LENGTH: 541
/ TYPE: PRT
/ ORGANISM: Vibrio cholerae
US-10-282-122A-77158

Alignment Scores: 1.34 Length: 541
Pctd. No.: 85.00 Matches: 51
Score: 51

```

```
Percent Similarity: 36.2% Conservative: 17
Best Local Similarity: 27.1% Mismatches: 62
Query Matchn: 9.1% Indels: 58
DB: 4 Gaps: 10
```

US-10-768-093-4 (1-527) x US-10-282-122A-77158 (1-541)

[illegible]

```

RESULT 10
US-10-369-493-2489
; Sequence 2489, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2489
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2489

Alignment Scores:      1.66      Length:      439
Pred. No.:

```



Score: 84.00 Matches: 43  
 Percent Similarity: 40.4% Conservatave: 24  
 Best Local Similarity: 25.9% Mismatches: 65  
 Query Match: 8.9% Indels: 34  
 Gaps: 9

US-10-768-093-4 (1-527) x US-10-369-493-2489 (1-439)

OY 54 TGGTGTATATGAGAAACAATTGCT-----TTAATTCTAATCTTGCTTCA 101  
 Db 131 TtrpYstrYrIleAlaGluAenPhelYThrLysGlyProTYrValLeuProValProser 150  
 OY 102 TTC-----GGCAGCCTATCCAGAACAGAAATAGCATTAAACTTCCCA 146  
 Db 151 PheAenValLeuAenGlySerHisAlaGlyAspLeuAlaPheGlnGluPheMet 170  
 OY 147 GTATCAACGACTATTTCAAAAGTTT----- 173  
 Db 171 IleuProThrGlyAlaProSerPheSerGluAlaMetArgTyrGlyAlaGluThrTyr 190  
 OY 174 -----TTGCACCTGAACCAATACAGCCTCTTTTGTTGAATGTTGA 221  
 Db 191 HisthrLeuLysSerIleAlaLysLysArgTyrGlySerSerAlaGly--AsnValGly 209  
 OY 222 AAGGAAGAGCTTATTATTATTAGTGTGAATTAACTGTTCTGAAATGTATCCAGGTA 281  
 Db 210 AspGluGlyGlyIle-----AlaProAspLeuGlnThrProGlnGluAlaLeuAspLeu 227  
 OY 282 ACGGTTCACCTGTTTATGATGAGATATTAGGTTAGACACACTAGTAAATACCCGTAT 341  
 Db 228 IleValGluAlaIleAenLysAlaGlyTyrGlu--GlyLysIleLysIleGlyLeuAsp 246  
 OY 342 GCTTCCCATCATATATCTACACATGTTGTATGAGAAAGGAAATATGTTAAAGAT 401  
 Db 247 ValAlaSerSerGluPheTyr-----ValAsp-----GlyLysTyrAspLeuAspIle 262  
 OY 402 CATGGTCAGAGGTTTACCATCATCAATCAATCACTTTAAAGCCGTGAAT-----TAT 455  
 Db 263 LysAlaAlaLysProLysProGluAsnLysLeuThrTyrGlnGlnLeuThrAspLeuTyr 282  
 OY 456 ACTAGCCGGGAAAAAAA 473  
 Db 283 ValGluLeuSerLysLys 288

## RESULT 11

US-10-267-989-25  
 ; Sequence 25, Application US/10267989  
 ; Publication No. US20030199035A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ARNAU, Jose  
 ; APPLICANT: VRANU, Asitrid  
 ; APPLICANT: ISRAELSEN, Hans  
 ; APPLICANT: JOERGENSEN, Flemming  
 ; APPLICANT: MADSEN, Soeren  
 ; TITLE OF INVENTION: METABOLICALLY ENGINEERED LACTIC ACID BACTERIA AND  
 ; TITLE OF INVENTION: MEANS FOR PROVIDING  
 ; FILE REFERENCE: ARNAU-1A  
 ; CURRENT APPLICATION NUMBER: US/10/267,989  
 ; PRIOR FILING DATE: 2002-10-10  
 ; PRIOR APPLICATION NUMBER: US/08/981,097  
 ; PRIOR FILING DATE: 1997-12-17  
 ; PRIOR APPLICATION NUMBER: PCT/DK97/00336  
 ; PRIOR FILING DATE: 1997-08-20  
 ; PRIOR APPLICATION NUMBER: 08/701,458  
 ; PRIOR FILING DATE: 1996-08-22  
 ; NUMBER OF SEQ ID NOS: 44  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 25  
 ; LENGTH: 740  
 ; TYPE: PRT  
 ; ORGANISM: Clostridium pasteurianum  
 ; US-10-267-989-25

## Alignment Scores:

Pred. No.: 1.91 Length: 740  
 Score: 84.00 Matches: 39  
 Percent Similarity: 38.6% Conservatave: 25  
 Best Local Similarity: 23.5% Mismatches: 68  
 Query Match: 8.9% Indels: 34  
 Gaps: 7

US-10-768-093-4 (1-527) x US-10-267-989-25 (1-740)

OY 93 CTTCGCTTCATTCGGCAGCCATGCCAGAAAGAAATAGCACTAAATCCAGTATCA 152  
 Db 326 IleAlaGlyValGlyIleAspGlyArgSerLeuValThrLysAenSerPheArgTyrLeu 345  
 OY 153 ACGACTATTTCAAAAAGTTTTCCTGACCTGACCAAGAAATACAGCCTTCTTTTGTCGA 212  
 Db 346 HisthrLeuIleAenLysSerAlaProGluProAsnMetThrValLeuTyrSerGlu 365  
 OY 213 AATGTTGAAAGGAGAGCTTTATTATTAGTGTGAATTAACTGTTCTGAAAAATGTA 272  
 Db 366 AsnLeu--ProGluSerPheLysPheCysAlaGluMetSerIleLeuThrAspSer 384  
 OY 273 TCCAGGTA-----ACGCTACCTGCTTTATGATGTAAGATTATGGTTA-- 317  
 Db 385 IleGlnTyrGluAenAspAspIleMetArgProIleTyrGlyAspAspTyrAlaIleAla 404  
 OY 318 -----GGACGACTAGTAAT 332  
 Db 405 CysCysValSerAlaMetArgValGlyLysAspMetGlnPhePheGlyAlaArgCysAsn 424  
 OY 333 ACCCGTATGCTTCCCATCATATCTACACATGTTGTATGAGAAAGGAAAAAATG 392  
 Db 425 LeuAlaLysCysLeuLeuAlaIleAenGlyLysAlaGlyLys-- 440  
 OY 393 TTAAAAATCATGTGTGACAGAGTTACACTTAAT--CAACAATATCTTTAAAGCGCTG 449  
 Db 441 -----LysGlyIleLysValValProAspIleGluProIleThrAspGluValLeu. 457  
 OY 450 AATTATACTAGCGGGGAAAAAATATATCTCTGGAATATATTAAGAT--CAGGTTATG 506  
 Db 458 AspTyr-----GluLysValLysGluAsnTyrPheLysValLeuGluTyrMet 473  
 OY 507 GTTGTTACTATGTAAAC 524  
 Db 474 AlaGlyLeuTyrValAsn 479

## RESULT 12

US-10-357-567-25  
 ; Sequence 25, Application US/10357567  
 ; Publication No. US20040038382A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Coschigano, Peter  
 ; TITLE OF INVENTION: Compositions and Methods for Bioremediation  
 ; FILE REFERENCE: OHU-07748  
 ; CURRENT APPLICATION NUMBER: US/10/357,567  
 ; PRIOR FILING DATE: 2003-02-04  
 ; PRIOR APPLICATION NUMBER: 09/072,433  
 ; PRIOR FILING DATE: 1998-05-04  
 ; PRIOR APPLICATION NUMBER: 60/046,845  
 ; PRIOR FILING DATE: 1997-05-05  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 25  
 ; LENGTH: 740  
 ; TYPE: PRT  
 ; ORGANISM: Clostridium pasteurianum  
 ; US-10-357-567-25

## Alignment Scores:

Pred. No.: 1.91 Length: 740  
 Score: 84.00 Matches: 39  
 Percent Similarity: 38.6% Conservatave: 25

Best Local Similarity: 23.5% Mismatches: 68  
 Query Match: 8.9% Indels: 34  
 DB: 4 Gaps: 7  
 US-10-768-093-4 (1-527) x US-10-357-567-25 (1-740)

OY 93 CTGGCTTATTCGGCAGCCATCCAGACAGAAATAGCGACTTAAAACTTCCCATATCA 152  
 DB 126 IleaAGlyValaIyIleapGlyArSerLeuValThrIleAAsenSerHearGlyLeu 345  
 OY 153 ACGACTATTTCAAAAGTTTTCGACCTGAACCAAGAAATACAGCTCTTTGGTGA 212.  
 DB 346 HistrleuIleAAsenGlySerIleAProGluProAAsenMetThrValleuSerGlu 365  
 OY 213 AATGTGGAAGAGAGCTTATTTATTTATGTCGAACTTCTCTGAAATGA 272  
 DB 366 AAsenLeu---ProGluSerPheIleGlySerPheCysAlaGluMetSerIleuThrAspSer 384  
 OY 273 TCCAGAGTA-----ACGCTTACCTGTTTATGATGAAGATTACGGTTA--- 317  
 DB 385 IlegInTyrgIuAAsenAAspIleMetArGProIleTyrgIyAAspApyrAlaIleAla 404  
 OY 318 -----CGACGACTAGTAAT 332  
 DB 405 CysCyValSerAlaMetArGValGlyLyAspMetGlnPhePheGlyAlaArgCysAsn 424  
 OY 333 ACCGCTGATGCTTCCCATCAATTAATCTACAGATTGTGATGAGAGGAAAAAATG 392  
 DB 425 LeuAlaIleCysLeuLeuLeuLeuAlaIleAAsnGlyValAAspGluLeu----- 440  
 OY 393 TTAAGATCATGTCGACAGAGTTACACCTAAT---CAACAAATTAATTAAAGCGCTG 449  
 DB 441 -----LyGlyIleLeuValIleProAAspIleGluProIleThrAspIleValleu 457  
 OY 450 AATTACTAGCGGGGAAAAAATATCTCTGAAATATATACGAT---CAGGTATG 506  
 DB 458 AAspTyT-----GluLySValLyGluAAsnTyPheLyValleuGluTyMet 473  
 OY 507 GTTGCTACTATGTAAC 524  
 DB 474 AlaGlyLeuTyValAAsn 479

RESULT 13  
 US-10-424-599-190153  
 ; Sequence 190153, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 190153  
 ; LENGTH: 157  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRTJ847\_142728C.1.pdp  
 US-10-424-599-190153

Alignment Scores:  
 Pred. No.: 1.45 Length: 157  
 Score: 83.50 Matches: 34  
 Percent Similarity: 35.3% Conservative: 19  
 Best Local Similarity: 22.7% Mismatches: 46  
 Indels: 51  
 Query Match: 8.9% Gaps: 5  
 DB: 4

US-10-768-093-4 (1-527) x US-10-424-599-190153 (1-157)  
 OY 81 TTAATTCATATTCCTTCATTCGCGCCATCCAGACAGAAATAGCGACTAAAAAC 140  
 DB 4 ValValThrIleCysThrSerGluSerAAsnArgAlaLyAspGluSerProThrAAsnGly 23  
 OY 141 TTCCAGATTCACAGACTATTTCAAAAGTTT----- 173  
 DB 24 PheLeuValleuSerLyIleGluAlaAspPheSerThrGlyLyTyrgIuIleSerGlu 43  
 OY 174 -----TTGACCTGAACCAAGAAATACAGCTTCTTTGGGAAATGT--- 218  
 DB 44 LeuValLySLeuIleAlaGluAAsnLyMetCysAlaGluAAsnHsValMetGluLySLeuVal 63  
 OY 219 -----GGAAGGAGAGCTTATTTATTTATGTCGAACTTAACTTACGTTCTGAA 266.  
 DB 64 GluAlaAspGluGlyLySgluAAspPheIleValTyrgIyValAAsnLeuThrPhe---Val 82  
 OY 267 AATGATCCAGAGTTACCGCTTAC----- 290  
 DB 83 AAsnLeuGluGluValAAsnIleTyAspGlyMetAAsnLeuAAsnGlyGlnTySProIleMet 102  
 OY 291 -----CCT 293  
 DB 103 AlaAsnCyThrPheArGlyValSerAAspLySglValValleuValleuProAlaPro 122  
 OY 294 GTTATGATGAAGATTATGGTTAGACGACTAGTAATACCGGTGATGCCATCA 353  
 DB 123 GluAAspArgIuAAspGlyGlyAAsnGlyArgIleVal---ThrValSerLeuProArgIu 141  
 OY 354 ATTAATCTACGAGTTGTGATGAGAAAGG 383  
 DB 142 GluLeuTyrgIleuLyAAspLySLeuGly 151

RESULT 14  
 US-10-424-599-250712  
 ; Sequence 250712, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 250712  
 ; LENGTH: 249  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRTJ847\_68422C.1.pdp  
 US-10-424-599-250712

Alignment Scores:  
 Pred. No.: 1.87 Length: 249  
 Score: 83.00 Matches: 38  
 Percent Similarity: 38.4% Conservative: 23  
 Best Local Similarity: 23.9% Mismatches: 56  
 Indels: 42  
 Query Match: 8.8% Gaps: 8  
 DB: 4

US-10-768-093-4 (1-527) x US-10-424-599-250712 (1-249)  
 OY 45 CAATTTATATGCTGTATATGAGAAAAATGTTTATTTCTTATTCCTTCATTC 104  
 DB 16 LysTyIleuTyrgCysTyT---HisArgThrIleuHisLyAAsnAlaProSerThrIlePhe 34  
 OY 105 GCGAGCCATGCCAGACAGAAATAGCG-----ACTAAAACTTCCGATTCACAGACT 158  
 DB: 4

Db 35 AsnAsnHsGlyThrAspGlyAsnAlaItrProThrSerLeuLeuProLeuAlaLeuAla 54  
Oy 159 ATTCAAAAAGTTTTTTTGCACTGACCAACCAATACACCTTTCTTT----- 206  
Db 55 ValSerIaGlySerLeuAlaLeuGlnProHisPheAsnProSerPheCysAspThrAsp 74  
Oy 207 -----GCTCAAAATGTTGGAAAGAGACCT----- 233  
Db 75 AspArgGlyValGlyValGlyGlySerThrGlnTyrValValGlySerGln 94  
Oy 234 -----TTATTTATTTAGTGAACCTTAACGTTCCCGAAAGATATCC 275  
Db 95 LysGluPheProAspGluLeuLeuGlnAspLeuLysIleValCysGlnAspAsnIleSer 114  
Oy 276 CAGTAAACGCTCTACCTGTTATGATGATGAGATTATGGTTAGCA----- 320  
Db 115 Leu-----AspTyrAspGluLysArgTyrIleHisGlyLysProGlnAsnSer 129  
Oy 321 -----CGACTAGTAATACCGCTGATGCTTCCCAATCAATATCTACAGATTGTTGAT 374  
Db 130 PheHisLysAlaValAsnIleProAspVal-----IleValTyrProArgSerGlu 146  
Oy 375 GAGAAAGGAAAAAATGTTAAA-----GATCATGGTGACAGAGTTACACCT 422  
Db 147 GluGluValSerLysIleValLysLeuCysAsnSerHisLysValProIleValPro 165

RESULT 15  
US-10-501-282-3988

/ Sequence 3988, Application US/10501282  
/ Publication No. US20050203280A1

/ GENERAL INFORMATION:

/ APPLICANT: MCMICHAEL, JOHN CALHOUN

/ APPLICANT: ZAGURSKY, ROBERT JOHN

/ APPLICANT: RUSSELL, DAVID PARRISH

/ APPLICANT: FLETCHER, LEAH DIANE

/ TITLE OF INVENTION: ALLOTOCOCUS OSTITIDIS OPEN READING FRAMES (ORFS) ENCODING  
/ TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF

/ FILE REFERENCE: AM100780 L2

/ CURRENT APPLICATION NUMBER: US/10/501,282

/ CURRENT FILING DATE: 2004-07-09

/ PRIOR APPLICATION NUMBER: 60/333,777

/ PRIOR FILING DATE: 2001-11-29

/ PRIOR APPLICATION NUMBER: 60/426,742

/ PRIOR FILING DATE: 2002-11-18

/ PRIOR APPLICATION NUMBER: PCT/US02/36123

/ PRIOR FILING DATE: 2002-11-25

/ NUMBER OF SEQ ID NOS: 6653

/ SOFTWARE: PatentIn version 3.2

/ SEQ ID NO 3988

/ LENGTH: 750

/ TYPE: PRT

/ ORGANISM: *Allolococcus ostitidis*

US-10-501-282-3988

Alignment Scores:

Pred. No.:	2.51	Length:	750
Score:	83.00	Matches:	38
Percent Similarity:	41.9%	Conservative:	32
Best Local Similarity:	22.8%	Mismatches:	61
Query Match:	8.8%	Indels:	36
DB:	5	Gaps:	8

US-10-768-093-4 (1-527) x US-10-501-282-3988 (1-750)

Oy 93 CTTCGCTTCATTCGGCAGCATGACGACAGAAAGAAATAGCAGTAATAAACTCCAGTATCA 152  
Db 335 ILeAlaGlyIleGlyLysAspGlyArgHisMetValThrLysAsnSerTyrArgPheLeu 354  
Oy 153 ACGACTATTTCAAAAGTTTTTTTGACACTGACCAACCAATACAGCCCTTTTGGTGA 212  
Db 355 HisThrLeuSerAsnLeuGlyProAlaIleProGluProAsnLeuThrValLeuIlePheSerGlu 374  
Oy 213 AATGTTGA-----AAGAAAGAGACCTTATATTATTAGTGAACCTTAACGTTCCT 263

Db 375 LysLeuProGlnAsnPheLysGluPheCysAlaLysValSerIleAlaSerSerAlaVal 394  
Oy 264 GAAATATATCCAGGTAAAGGCTCTACCTGTTATGATGATGATTAAGGTTAGACGA 323  
Db 395 GlnTyrGluAsnAspGluIleMetArgProTyrPheGlyAspAspTyrGlyIleAlaCys 414  
Oy 324 CTAGTAATATCC----- 344  
Db 415 CysValSerLysMetGluIleGlyLysGlnMetGlnPhePheGlyAlaArgAlaAsnLeu 434  
Oy 345 TCCCAATCAATATCTACAGATT-----GTGATGAGAAAGGAAAAAATGTGA 395  
Db 435 AlaLysAlaLeuLeuTyrAlaIleAsnGlyGlyValAspGlu-----ValThr 450  
Oy 396 AAGATCATGATGTCAGAGAGTTACACCTAAT---CAACAAATTAATTTAAAGCCGTGAAT 452  
Db 451 LysAsp-----GlnValGlyProLysTyrGluProIleThrSerGluTyrLeuAsp 467  
Oy 453 TATPCTAGCGGGGAAAAAATATCTCTCGAATATATATACGATCGATT----- 503  
Db 468 TyrAspGluValValGluLysTyr-----AspGlnMetThrGluTyr 481  
Oy 504 ATGTTGTTACTATGTAAC 524  
Db 482 LeuAlaGlyLeuTyrIleAsn 488

Search completed: April 10, 2006, 10:12:53  
Job time : 129.5 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: April 10, 2006, 10:08:49 ; Search time 3.4 Seconds  
(without alignments)  
966.951 Million cell updates/sec

Title: US-10-768-093-4  
Sequence: 1 TTGCACATTCAGATGTTA.....TTGGTACTACTGTAACATA 527

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 184161 seqs, 31191982 residues  
Total number of hits satisfying chosen parameters: 368322

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p model -DEV=xlp  
-O=abs/ABSWEB\_pool/US10768093/runat\_10042006\_090311\_4719/app\_query.fasta.1  
-DB=Published\_Applications\_AA\_New-OPM=faaban-SUFFIX=n2p.rapbn-MINMATCH=0.1  
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -HOST=absn2p  
-USER=US10768093@CGN1.1.25@runat\_10042006\_090311\_4719 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA\_New:  
1: /SIDSS/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp:\*  
2: /SIDSS/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp:\*  
3: /SIDSS/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp:\*  
4: /SIDSS/ptodata/2/pubpaa/PCR\_NEW\_PUB.pcp:\*  
5: /SIDSS/ptodata/2/pubpaa/US05\_NEW\_PUB.pcp:\*  
6: /SIDSS/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp:\*  
7: /SIDSS/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp:\*  
8: /SIDSS/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90.5	9.6	268	US-11-195-739-6	Sequence 6, Appl1
2	80.5	8.6	379	US-11-096-568A-6312	Sequence 6312, Ap
3	80.5	8.6	416	US-11-096-568A-6311	Sequence 6311, Ap
4	80.5	8.6	450	US-11-096-568A-6310	Sequence 6310, Ap
5	78.5	8.4	1198	US-10-451-375-4	Sequence 4, Appl
6	77.5	8.3	1162	US-10-451-375-3	Sequence 3, Appl1
7	75.5	8.0	356	US-11-096-568A-32647	Sequence 32647, A
8	75.5	8.0	361	US-11-096-568A-32646	Sequence 32646, A
9	75.5	8.0	1680	US-10-517-939-362	Sequence 362, App

10	74.5	7.9	400	US-10-793-626-2774	Sequence 2774, Ap
11	73.5	7.8	273	US-11-051-720-1375	Sequence 1375, Ap
12	73.5	7.8	729	US-11-051-720-1373	Sequence 1373, Ap
13	73.5	7.8	738	US-11-051-720-1374	Sequence 1374, Ap
14	73.5	7.8	1006	US-10-793-626-154	Sequence 154, App
15	73.5	7.8	1081	US-11-051-720-1372	Sequence 1372, Ap
16	73.5	7.8	1767	US-10-995-561-911	Sequence 911, App
17	73.5	7.8	1767	US-10-995-561-914	Sequence 914, App
18	73.5	7.8	1806	US-10-995-561-912	Sequence 912, App
19	73.5	7.8	1806	US-10-995-561-915	Sequence 915, App
20	73.5	7.8	1806	US-11-051-720-1446	Sequence 1446, Ap
21	73.5	7.8	1806	US-11-051-720-1447	Sequence 1447, Ap
22	73.5	7.8	1818	US-10-995-561-917	Sequence 917, App
23	73.5	7.8	1818	US-10-995-561-913	Sequence 913, App
24	72.5	7.7	375	US-10-485-517-166	Sequence 166, App
25	72.5	7.7	465	US-10-873-528-164	Sequence 164, App
26	70.5	7.5	357	US-09-995-493-178	Sequence 178, App
27	70.5	7.5	357	US-11-194-246-282	Sequence 282, App
28	70.5	7.5	501	US-11-087-099-3301	Sequence 3301, Ap
29	70.5	7.5	525	US-10-763-712A-108	Sequence 108, App
30	70.5	7.5	898	US-11-099-691-7	Sequence 7, App1
31	70.5	7.5	898	US-11-124-367A-330	Sequence 330, App
32	70.5	7.5	898	US-11-124-367A-333	Sequence 333, App
33	70.5	7.5	905	US-11-087-099-433	Sequence 433, App
34	70	7.5	321	US-11-096-568A-7764	Sequence 7764, Ap
35	70	7.5	334	US-11-096-568A-7763	Sequence 7763, Ap
36	70	7.5	361	US-11-096-568A-7762	Sequence 7762, Ap
37	69.5	7.4	505	US-11-087-099-5784	Sequence 5784, Ap
38	69.5	7.4	501	US-11-087-099-8619	Sequence 8619, Ap
39	69	7.3	209	US-11-096-568A-1094	Sequence 1094, Ap
40	69	7.3	249	US-11-096-568A-1093	Sequence 1093, Ap
41	69	7.3	1432	US-10-510-386-218	Sequence 218, App
42	68.5	7.3	278	US-10-793-626-2208	Sequence 2208, App
43	68.5	7.3	340	US-10-506-454-335	Sequence 335, App
44	68.5	7.3	328	US-11-057-012-52	Sequence 52, App1
45	68.5	7.3	484	US-11-087-099-1387	Sequence 1387, Ap

ALIGNMENTS

RESULT 1  
US-11-195-739-6  
; Sequence 6, Application US/11195739  
; Publication No. US20060019324A1  
; GENERAL INFORMATION:  
; APPLICANT: ALZARI, PEDRO  
; APPLICANT: BOITEL, BIRGITTE  
; APPLICANT: VILLARINO, ANDREA  
; APPLICANT: FERNANDEZ, PABLO  
; APPLICANT: COLE, STEWART  
; TITLE OF INVENTION: PKB KINASE AND PSTP PHOSPHATASE AND METHODS OF IDENTIFYING  
; FILE REFERENCE: 252853US  
; CURRENT APPLICATION NUMBER: US/11/195,739  
; CURRENT FILING DATE: 2005-08-03  
; PRIOR APPLICATION NUMBER: US/10/892,170  
; PRIOR FILING DATE: 2004-07-16  
; PRIOR APPLICATION NUMBER: US 60/487,943  
; PRIOR FILING DATE: 2003-07-18  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patencin version 3.2  
; SEQ ID NO 6  
; LENGTH: 268  
; TYPE: PRT  
; ORGANISM: Thermobifida fusca  
US-11-195-739-6  
Alignment Scores: 0.0625  
Pred. No.: 90.50  
Percent Similarity: 42.1%  
Score: 42.1%  
Best Local Similarity: 26.8%  
Query Match: 9.6%  
Length: 268  
Matches: 49  
Conservative: 28  
Mismatch: 54  
Indels: 52

[illegible]

Alignment Scores:	
Pred. No.:	0.836
score:	80.50
Percent Similarity:	34.0%
Best Local Similarity:	22.0%
Query Match:	6.6%
DB:	7
Length:	379
Matches:	33
Conservative:	18
Mismatches:	48
Indels:	51
Gaps:	5

US-10-768-093-4 (1-527) x US-11-096-568A-6312 (1-379)

QY 81 TTAATTCATTAATTCCTTGCTTCATTCGGCAGCCATCCGACGAAACAGAAAATAGCGCTAAAAAC 140  
Db 226 ValValThrTlleCysThrSerGluSerAsnArgAlaGluAsnGluPheProThrAsnGly 245  
QY 141 TTCGCCGATTCACACGACTATTTCACAAAAGTTT----- 173  
Db 246 PheLeuValLeuSerLeuIleGluAlaAspPheSerThrGlyLeuTyrgluIleSerGlu 265  
QY 174 -----TTTGCCACTGAAACACGAAATACAGCTCTTTTGTTGTAATAATGTT--- 218  
Db 266 LeuValLySerLeuIleAlaGluAsnLysMetValGluIleAsnHisValMetGluLysLeuVal 285  
QY 219 -----GGAAAGGAAGAGACTTATTATTATTTAGTGCAACTTAATCTGTTCTCGAA 266  
Db 286 GluAlaAspGluGlyLysGluAspPheIleValTyrglyValAsnLeuThrPhe---Val 304  
QY 267 AATGATATCCAGGTAAACGGTCTAC----- 290  
Db 305 AsnLeuGluGluAlaAsnIleTyraAspGlyMetAsnLeuAsnGlyGlnLysProIleMet 324  
QY 291 -----CCT 293  
Db 325 AlaAsnCysThrPheAsnGlyValSerAspLysGlyValValLeuValLeuPheProAlaPhe 344  
QY 294 GTTTATGATGAGATTAATGAGGTGTAAGACGACTAGTAATAATACCGCTGATGCTTCCCATCA 353  
Db 345 GluAspAspGluAspGlyGlyAsnGlyArgGluIleVal---ThrValSerLeuProArgGlu 363  
QY 354 ATATATCTACCAAGATTGTTGATGACAAAAGG 363  
Db 364 GluLeuTyrgluIleuLysAspLysLeuGly 373

```

RESULT 3
US-11-096-568A-6311
? Sequence 6311, Application US/11096568A
? Publication No. US20060048240A1
? GENERAL INFORMATION:
? APPLICANT: Alexandrov, Nikolai et al.
? TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
? TITLE OF INVENTION: Thebby
? FILE REFERENCE: 2750-1592PUS2
? CURRENT APPLICATION NUMBER: US/11/096,568A
? CURRENT FILING DATE: 2005-04-01
? NUMBER OF SEQ ID NOS: 34471
? SEQ ID NO 6311
? LENGTH: 416
? TYPE: PRT
? ORGANISM: Glycine max
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)..(416)
? OTHER INFORMATION: Cereus Seg. ID no. 14314719
US-11-096-568A-6311

Alignment Scores:
Pred. No.:          0.851          Length:          416
Score:              80.50          Matches:         33
Percent Similarity: 34.0%          Conservative:    18
Best Local Similarity: 22.0%       Mismatches:     48
Query Match:        8.6%           Indels:         51
DB:                 7              Gaps:           5
US-10-768-093-4 (1-527) x US-11-096-568A-6311 (1-416)

QY      81  TTAATTCTAATTCCTTGCTCATTCGGAGCCAGCCAGAAACAAATATGCACTTAAATAAC 140
      ::::  |||  :::  |||
Db      263  VAIVAlhrIleCyGtnrSergIuserGlsrAphgIadIuhengIuPheProThrsmcIy 282
      |||  |||  :::  |||

QY      141  TTCCCGATATCAAGCACTATTCTCAAAAAGTTT----- 173
      |||  |||  :::  |||

```

```

Db      263 PheLeuValIleuSerLysIleGluIaAspHseSerThrGlyLysTrpGluIleSerGlu 302
Oy      174 -----TTTGCACCTGAAACCAACGATACAGCCCTTTGGTGAATAAGTT--- 218
Db      303 LeuValLysLeuIleAlaGluIuAsnLysMetValGluAsnHleValMetGluLysLeuVal 322
Oy      219 -----GGAAGGAGGAGCGCTTATTATTAGTGAACCTAATCTGTCTGAA 266
Db      323 GluIaAspGluGlyLysGluAspPheIleValTrpGlyValAsnLeuThrPhe---Val 341
Oy      267 AATGATCCAGCATGAAACGGCTTAC----- 290
Db      342 AsnLeuGluGluAlaAsnIleTrpAspGlyMetAsnLeuAsnGlyLysProIleMet 361
Oy      291 ----- 293
Db      362 AlaAsnGlyThrPheArgGlyValSerAspLysGlyValIleValLeuValLeuProAlaPro 381
Oy      294 GTTATGATGTAAGATTATGCGCTTAGAGACAGCATAGTAAATACCGGTGATGCTCCCAATCA 353
Db      382 GluAspAspGluAspGlyGlyAsnGlyAsnIleVal---ThrValSerLeuProArgGlu 400
Oy      354 ATAATCTACAGATGTTGATGATGAAGG 383
Db      401 GluLeuTrpGlnLeuLysPheLysLeuGly 410

RESULT 4
US-11-096-568A-6310
; Sequence 6310, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theory
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6310
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(450)
; OTHER INFORMATION: Ceres Seq. ID no. 14314718
US-11-096-568A-6310

Alignment Scores:
Pred. No.: 0.865 Length: 450
Score: 80.50 Matches: 33
Percent Similarity: 34.0% Conservative: 18
Best Local Similarity: 22.0% Mismatches: 48
Query Match: 8.6% Indels: 51
DB: Gaps: 5

US-10-768-093-4 (1-527) x US-11-096-568A-6310 (1-450)
Oy      81 TTAATTCATATTCCTTCCTTCATTCGCGACGCCATGCCAGAACGAATAAGCACTAATAAC 140
Db      297 ValValThrIleCysThrSerGluSerAsnArgAlaGluAsnGluPheProThrAsnGly 316
Oy      141 TTCCAGATCAACAGCATATTCCTCAAAAAGTTT----- 173
Db      317 PheLeuValLeuSerLysIleGluIaAspHseSerThrGlyLysTrpGluIleSerGlu 336
Oy      174 -----TTGCACTGAAACCAAGATACAGCCCTTTGGTGAATAAGTT--- 218
Db      337 LeuValLysLeuIleAlaGluAsnLysMetValGluAsnHleValMetGluLysLeuVal 356
Oy      219 -----GGAAGGAGGAGCGCTTATTATTAGTGAACCTAATCTGTCTGAA 266
Db      357 GluIaAspGluGlyLysGluAspPheIleValTrpGlyValAsnLeuThrPhe---Val 375

```

```

QY 261 AATGATATCCAGTAAAGCTTAC----- 260
      |||:::|||||
Db 376 AemLEGLUGLAlaAmLEtYrAspGLYMeLsAmLEuBsnGLYGLnLYProLIeMeC 355
      |||:::|||||
QY 291 -----CCT 293
Db 396 AlaAenCYsThrPheArGGLYValSerAspLYsGLYValLeuValLeuProLAlaPro 415
      |||:::|||||
QY 294 GTTTATGATGAAGATTATGGCTTAGACGACTAGTAATACCGGTATGCTCCCAATCA 353
      |||:::|||||
Db 416 GLuAspAspGLuAspGLYGLYAsnGLYArGLIleVal---ThrValSerLeuProArGLu 434
      |||:::|||||
QY 354 ATAATCTACACAGATTGTTGATGAAGAAAGG 383
      |||:::|||||
Db 435 GLuLeuTYrGLInLeuLYAspLYLeuSLY 444
      |||:::|||||

RESULT 5
US-10-451-375-4      / Sequence 4: Application US/10451375
US-10-451-375-4      / Publication No. US20050261482A1
GENERAL INFORMATION:
APPLICANT: Bayer AG
TITLE OF INVENTION: REGULATION OF HUMAN SERINE-THREONINE PROTEIN KINASE
FILE REFERENCE: L10264 Foreign Countries
CURRENT APPLICATION NUMBER: US/10/451,375
PRIOR FILING DATE: 2003-07-02
PRIOR APPLICATION NUMBER: US 60/259,215
PRIOR FILING DATE: 2001-01-03
PRIOR APPLICATION NUMBER: US 60/306,468
PRIOR FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: US 60/308,098
PRIOR FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
      / SEQ ID NO 4
      / LENGTH: 1198
      / TYPE: PRT
      / ORGANISM: Homo sapiens
US-10-451-375-4

Alignment Scores:
Pred. No.:      1.75      Length:      1198
Score:          78.50     Matches:      39
Percent Similarity: 42.0%  Conservative: 24
Best Local Similarity: 26.0%  Mismatches: 64
Query Match:      8.4%     Indels:      23
DB:              6        Gaps:      8
US-10-768-093-4 (1-527) x US-10-451-375-4 (1-1198)

QY 93 CTTCGCTTACATTCGGCAGCCATCCAGACAACGAATAATACGACTAATAAATTTC----- 143
      |||:::|||||
Db 41 MetThrGLYTYrGLYSerHisSerHisValTYrSerGLnSerLYsAsnLEProSer 60
      |||:::|||||
QY 144 ---CCAAGTATCAACGACATATTCAAAAAGTTTTCGACCTGACCAACGACATACAG--- 197
      |||:::|||||
Db 61 GLnProLAlaSerThrThrValSerThrSerLeuProValProAsnProSerLeuProTYr 80
      |||:::|||||
QY 198 -----CCTTCTTTTGGTGAAAATGTTGAAAAGGAAGAGGCTTTATTA 239
      |||:::|||||
Db 81 GLInGlnThrIleValPheProGLYSerThrGLYHisIleValValThrSerAlaSerSer 100
      |||:::|||||
QY 240 TTTATGTGTGAAC-----TTTACTGTTCTCGAATAATGTA-----TCCCAAGTA 281
      |||:::|||||
Db 101 ThrSerValThrGLYGLnValLeuGLYGLYProHisLeuLeuSerArgArGLSerThrVal 120
      |||:::|||||
QY 282 ACGGCTACCCCTGTTTATGATGAAGATTATGGCTTAGACGACTAGTAATACCGCTGAT 341
      |||:::|||||
Db 121 SerLeuLeuAspThrTYr---GLnLYsCYsGLYLeuLYsArGLYSerGLuGLuIleGLu 139
      |||:::|||||
QY 342 GCTTCCCAATGAATATATCTACCAAGATTGTGATGAGAAAGGAAAAAATGTTAAAGAT 401
      |||:::|||||

```

Db 140 AenThSerSeVal-----GlnIleIleGluGluHis---ProPronectIleGlnAsn 156  
 Oy 402 CATGTGTGACAGGTTACACCTTAATCAACAATAATCTTTAAAGCGTGAATTACTAGC 461  
 Db 157 AenAlaSerGlyAlaThrValAlaThrAlaThrThSerThrAlaThrSerLysAsnSer 176  
 Oy 462 GGGGAAAAAAATATCTCTCGAATATAT 491  
 Db 177 GlySerAsn-----SerGluGlyAspTyr 184

RESULT 6  
 US-10-451-375-3  
 / Sequence 3, Application US/10451375  
 / Publication No. US20050261482A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Bayer AG  
 / TITLE OF INVENTION: REGULATION OF HUMAN SERINE-THREONINE PROTEIN KINASE  
 / FILE REFERENCE: LIO264 Foreign Countries  
 / CURRENT APPLICATION NUMBER: US/10/451,375  
 / CURRENT FILING DATE: 2003-07-02  
 / PRIOR APPLICATION NUMBER: US 60/259,215  
 / PRIOR FILING DATE: 2001-01-03  
 / PRIOR APPLICATION NUMBER: US 60/306,468  
 / PRIOR FILING DATE: 2001-07-20  
 / PRIOR APPLICATION NUMBER: US 60/308,098  
 / NUMBER OF SEQ ID NOS: 14  
 / SOFTWARE: PatentIn version 3.1  
 / SEQ ID NO 3  
 / LENGTH: 1162  
 / TYPE: PRT  
 / ORGANISM: Mesocricetus auratus  
 / US-10-451-375-3

Alignment Scores:  
 Pred. No.: 2.23 Length: 1162  
 Score: 77.50 Matches: 40  
 Percent Similarity: 42.0% Conservative: 23  
 Best Local Similarity: 26.7% Mismatches: 64  
 Query Match: 8.3% Indels: 23  
 Gaps: 8

US-10-768-093-4 (1-527) x US-10-451-375-3 (1-1162)

Oy 93 CTGTGCTTATTCGCGACCCATGCCAGAACAGAAATAGCGACTTAATCTTC----- 143  
 Db 34 MetThGlyTyrGlySerHisSerLysValTyrSerGlnSerLysAsnIleProProSer 53  
 Oy 144 ---CCGATATCAAGCATATTTTCAAAAGTTTTCGACCTGACCA----- 191  
 Db 54 GlnProAlaSerThrThrValSerThrSerLeuProIleProAsnProSerLeuProTyr 73  
 Oy 192 -----ATACAGCCTTCTTTTGTGAAAGTTTGGAAGAGAGAGAGCTTATTA 239  
 Db 74 GluGlnThrIleIleLeuProGlySerThrGlyHisIleValValThrSerAlaSerSer 93  
 Oy 240 TTATAGTGTGAAC-----TTACTGTTCTGAAATGTA-----TCCAGGTA 281  
 Db 94 ThrSerValThrGlyGlnValLeuGlyGlyProHisAsnLeuMetArgatGSerThrVal 113  
 Oy 282 ACCGCTTACCTTTTATGATGTAAGATATAGGTTAGAGACGACTAGTAATACCGGTAT 341  
 Db 114 SerLeuLeuAspThrTyr---GlnLysCybGlyLeuLysArgLysSerGluGluIleGlu 132  
 Oy 342 GCTTCCCAATCAATTAATCTACAGATGTTGATGAGAAAGGAAAAAATGTTAAAAAGAT 401  
 Db 133 AenThSerSerVal-----GlnIleIleGluGluHis---ProProMetIleGlnAsn 149  
 Oy 402 CATGTGTGACAGGTTACACCTTAATCAACAATAATCTTTAAAGCGTGAATTACTAGC 461  
 Db 150 AenAlaSerGlyAlaThrValAlaThrAlaThrThSerThrAlaThrSerLysAsnSer 169  
 Oy 462 GGGGAAAAAAATATCTCTCGAATATAT 491

Db 170 GlySerAsn-----SerGluGlyAspTyr 177

RESULT 7  
 US-11-096-568A-32647  
 / Sequence 32647, Application US/11096568A  
 / Publication No. US20060048240A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Alexandrov, Nikolai et al.  
 / TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 / FILE REFERENCE: 2750-1592PUS2  
 / CURRENT APPLICATION NUMBER: US/11/096,568A  
 / CURRENT FILING DATE: 2005-04-01  
 / NUMBER OF SEQ ID NOS: 34471  
 / SEQ ID NO 32647  
 / LENGTH: 356  
 / TYPE: PRT  
 / ORGANISM: Arabidopsis thaliana  
 / FEATURE:  
 / NAME/KEY: misc\_feature  
 / LOCATION: (1)..(356)  
 / OTHER INFORMATION: Ceres Seq. ID no. 13593732

US-11-096-568A-32647

Alignment Scores:  
 Pred. No.: 2.91 Length: 356  
 Score: 75.50 Matches: 28  
 Percent Similarity: 44.0% Conservative: 16  
 Best Local Similarity: 28.0% Mismatches: 23  
 Query Match: 8.0% Indels: 33  
 Gaps: 6

US-10-768-093-4 (1-527) x US-11-096-568A-32647 (1-356)

Oy 258 GTTCCGAAATGTATCCAGGTA-----ACGGTACCTGTTAT 299  
 Db 23 IleProGlnAsnArgAlaGlnIleProGlnTyrPheLysThrAsnValLysProTyrSer 42  
 Oy 300 GATGAAGATTATGGTTAGGACGACTAGTAATACCGCTGATGCTTCCCAATCAATATC 359  
 Db 43 GlnArgLysGlyThrLeuAspProAlaLeuGlnAlaGlnAlaAlaArgGlnIleIle 62  
 Oy 360 TACCAATTTGTGACGAAAGGAAAAAATGTTAAAGATCAATGCTGACAGGTTTACA 419  
 Db 63 -----ThrValAsnGlnLysGly-----GlyVala----- 70  
 Oy 420 CCTAATCAACAATATCTTTAAAGCGTGAAT-----TATACTACCGG 464  
 Db 71 -----AsnPheLysThrLeuAsnGlnValIleLysSerIleProThrGly 85  
 Oy 465 GAAAA-----AAATATCTCTCGAATATATATACGATGATGTTATGTT 509  
 Db 86 AsnLysAsnArgValIleIleLysLeuAlaProGlyValLysAsnGlnLysValThrIle 105

RESULT 8  
 US-11-096-568A-32646  
 / Sequence 32646, Application US/11096568A  
 / Publication No. US20060048240A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Alexandrov, Nikolai et al.  
 / TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
 / FILE REFERENCE: 2750-1592PUS2  
 / CURRENT APPLICATION NUMBER: US/11/096,568A  
 / CURRENT FILING DATE: 2005-04-01  
 / NUMBER OF SEQ ID NOS: 34471  
 / SEQ ID NO 32646  
 / LENGTH: 361  
 / TYPE: PRT  
 / ORGANISM: Arabidopsis thaliana  
 / FEATURE:  
 / NAME/KEY: misc\_feature









